

OM of: US-09-301-906-3 to: A\_Geneseq\_36:\* out\_format : pfs  
Date: Jan 31, 2001 3:40 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=frame+ntp.model -DEV=xlp  
-O/cg2\_1/JUSPTO\_spool/US09301906/runat\_31012001\_153034\_9577/app\_query.fasta\_1.6881  
-DB=A\_Geneseq\_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.500 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blisum62  
-TRANS-human40.cdl -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=EXT -MINLEN=0 -MAXLEN=200000000  
-USER=US09301906\_@CGN1\_1\_396 -ICPU=6 -NCPU=6 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:  
Query: US-09-301-906-3  
Query length: 6714  
Database: A\_Geneseq\_36:\*  
Database sequences: 268485  
Database length: 34193795  
Search time (sec): 277.730000

score\_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58148			11520.00	19619.90	..
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:W21636			5247.00	8908.82	0.0
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58145			1495.00	2520.59	3.1e-133
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58144			1432.00	2413.92	2.9e-127
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W73476			802.00	1305.05	1.8e-66
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y91969			631.50	1044.38	4.9e-51
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58143			605.00	1012.62	7.5e-49
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y91963			323.00	507.85	2.0e-21
/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:R05107			323.00	498.79	3.3e-21
/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:R06068			309.00	475.02	7.1e-20
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W87724			230.00	433.12	7.9e-18
/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:R06531			274.50	417.76	1.2e-16
/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:R49657			237.50	346.81	6.4e-13
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:Y31381			186.50	259.72	4.6e-08
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:Y31384			185.00	257.08	6.4e-08
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W56322			185.00	254.55	7.4e-08
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W6368			184.50	256.35	7.1e-08
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W93408			181.00	250.24	1.5e-07
/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:R51284			179.50	247.81	2.1e-07
/SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:R91813			179.50	247.81	2.1e-07
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W61519			179.50	247.81	2.1e-07
/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:R05562			179.00	248.33	2.2e-07
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W80196			169.50	230.72	1.9e-06
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W71209			169.50	230.72	1.9e-06
/SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:W03740			160.50	217.25	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W95160			160.50	217.25	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W89299			160.50	217.25	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y78508			160.50	217.25	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y68447			160.50	217.25	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:R14618			160.50	215.34	1.4e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:W06725			158.50	221.79	1.2e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W47273			158.00	209.08	2.6e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:P93357			154.50	206.61	4.6e-05
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y59288			152.50	201.57	7.9e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:R08031			151.50	208.12	6.1e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:R05041			150.00	199.86	9.8e-05
/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:R47578			149.50	205.67	8.9e-05
/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y87407			147.00	181.51	0.0005

/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:W88417 + 146.50 200.01 0.0002 928  
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:Y35060 + 146.50 199.69 0.0002 949  
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:Y37825 + 146.50 196.41 0.0002 1194  
/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:Y05477 - 145.00 200.49 0.0002 750  
/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:W41935 + 145.00 193.39 0.0003 1233

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58148

seq\_documentation\_block:

ID	Y58148	standard; Protein: 2237	AA.
AC	Y58148;		
DT	07-MAR-2000	(first entry)	
DE	GLRaV-3	polyprotein.	
KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;		
KW	viral disease; yield loss; sugar content; inhibition; infection;		
KW	replication; polyprotein; domain; proteinase; methyltransferase;		
KW	helicase; RNA-dependent; RNA polymerase; untranslated region;		
KW	transgenic plant; component; resistant; truncation; deletion; antisense;		
KW	expression; detection; antibody.		
OS	Grapevine leafroll-associated virus 3.		
XX			
PH	Key	Location/Qualifiers	
FT	Domain	85..204	
FT		/note= "Proteinase domain (Y58143)"	
FT	Domain	460..731	
FT		/note= "Methyltransferase domain (Y58144)"	
FT	Domain	1922..2212	
FT		/note= "Helicase domain (Y58145)"	
XX			
PN	W09955880-A1.		
XX			
PD	04-NOV-1999.		
XX			
PF	29-APR-1999;	99WO-US09307.	
XX			
PR	29-APR-1998;	98US-0083404.	
XX			
PA	(CORR ) CORNELL RES FOUND INC.		
XX			
PI	Gonsalves D, Ling K;		
XX			
DR	WPI: 2000-062035/05.		
XX			
DR	N-PSDB; Z49208.		
XX			
PT	Newly isolated grapevine leafroll virus protein or polypeptide useful		
PT	for producing transgenic plants conferring viral disease resistance -		
XX			
PS	Claim 3; Fig 16; 84pp; English.		
XX			
CC	This sequence represents the polyprotein encoded by open reading		
CC	frame (ORF) 1a from the grapevine leafroll-associated virus 3		
CC	(GLRaV-3). Leafroll is a serious viral disease, occurring wherever		
CC	grapes are grown. Although not lethal, it causes yield losses and		
CC	reduction in sugar content. The virus encodes several proteins,		
CC	which may serve as targets for the inhibition of viral infection or		
CC	replication. These proteins include the 242-248 kD polyprotein (Y58148),		
CC	encoded by open reading frame (ORF) 1a, and which comprises a proteinase		
CC	domain (Y58143), a methyltransferase domain (Y58144) and a helicase		
CC	domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF		
CC	1b; and a protein of unspecified function (Y58147), encoded by ORF 11.		
CC	Nucleotides encoding these proteins, or fragments thereof, and the 5'		
CC	and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are		
CC	useful for the generation of transgenic plants and plant components.		
CC	Such transgenic plants may be resistant to viral disease, for example,		
CC	this property being conferred on the plants via the use of nucleotides		
CC	encoding truncated or internally deleted proteins, or via the use of		
CC	antisense nucleotides to inhibit viral gene expression. The nucleotides		
CC	may additionally be used in the detection of viral nucleic acids in a		

CC tissue sample. The proteins can be used to generate antibodies which  
CC may be used to detect GURAV-3 in plant samples. The isolation of  
CC GURAV-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.  
XX

50 Sequence 2237 AA;

alignment\_scores:  
Quality: 11520.00 Length: 2237  
Ratio: 5.150 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-301-906-3 x Y58148 ..

Align seg 1/1 to: Y58148 from: 1 to: 2237

```
1  A'GGACTACATTCGGCCCATTTGGCGCTTTTCTCCTTTCTCAGCTTAAATAA 50
|||||
1  MetAspTyrIleArgProLeuArgValPheSerPheProHisValAsnAs 17
51  CACCTTGGAGTACGTTAGGTACAAAGCCCAATGGTGATGTAGGAGCTT 100
|||||
17  nThrLeuGluTyrValAa-gTyrAsnLysAlaAsnGlyAspValGlyAlaP 34
|||||
101 TCCTAACCAACCATGAAGTTCATAGGAACGCTGAAGTTGTGCGACTTCACA 150
|||||
34  heLeuThrThrMetLysPheIleGlyAsnValLysLeuSerAspPheThr 50
|||||
151 CCCAGTGGCAGCTATGATTTACATTTGAAAGCTTCACCAAGGGGTGAA 200
|||||
51  ProArgCysAlaAlaMetIleTyrIleGlyLysLeuThrLysGlyVally 67
201 CGGTACGTTTGTCCCCACACAGTTAAAGGGTTTGCACGGCAGTAGCGTG 250
|||||
67  sargThrPheValProProProValLysGlyPheAlaArgGlnTyrAlaIa 84
|||||
251 TTGTCCAGCGGCTCAGTCAGCGCGCTGAGAGGGGATGGTAAGAAGGCTTG 300
|||||
84  alValSerGlySerValSerAlaLeuAa-gLyspGlyLysLysValleu 100
|||||
301 A'GGAGGCAAGGACCTCAACTTCCGCAACTTCCGACGCTGCTGTGATTTCCA 350
|||||
101 MetGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAs 117
|||||
351 CGTCGTATTCCGAAGCTGTTCTAATGCAATTACTTTCGTACACTACCACC 400
|||||
117 pValValPheGluAlaValSerAsnAlaLeuLeuValValHisTyrHisA 134
|||||
401 GGGTAGTGCCTGATGCCCGCTCAAGCGCGAGCAGCCTAAACCGGCTGTT 450
|||||
134 qValValProTyrAlaProValLysArgGluGlnProLysProAlaVal 150
|||||
451 AAGCAAGATGACGACGAAGCCCAACGGAAGCGTCAATTTGGCGTGTAA 500
|||||
151 LysGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaVally 167
|||||
501 GCCACACGCTGTTGGCGTCCACGCTACCCTTCCCTAAACCAAGGAGCAC 550
|||||
167 sProThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaL 184
|||||
551 TGGAGCCACGCGCAATCAGTCCCACAAACAGTCGTTGGAGGGAAGGCCGCC 600
|||||
184 euGluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAla 200
|||||
601 TTGACGTTTGGCCTTTCTCAGTAAGGCTGGGCTGATGAGAGCCACGC 650
|||||
201 LeuThrPheGlyLeuPhePheSerLysGlyGlyAspGluSerAspAl 217
|||||
651 TGTCACTTTCGGGAAGGGAATTTGTTTAACAGGGCCCTTATGTTCCCTA 700
|||||
217 aValIleLeuArgLysGlyLysLeuPheAsnArgAlaLeuAsnValProI 234
|||||
```



1601 TCCTAGACGTTAAAGATGTCGACGGAGAAATCAATGAGATCCCTCTTCTT 1650  
534 allLeuAspValLysAspValLysArgArgIleAsnGluIleLeuPheLeu 550  
1651 TCCACAGCTGGGGAGATTCGTAGCTGCCAGTGACCTTCTAAGTGAAGC 1700  
551 SerThrAlaGlyGlyAspSerTyrValSerSerAspLeuLeuThrGluAl 567  
1701 GGCTTCAAGCTGTGCTTACTGTAGTCGAGATCCGAGAACTGCAGATT 1750  
567 aAlaSerLysSerValSerTyrCysSerArgGluSerGlnAsnCysAsps 584  
1751 CTAGAGCCGATCGGGTTTATGGTGGATCTGTACGATATATCCCGCAG 1800  
584 erArgAlaaspAlaGlyPheMetValaspValTyrAspIleSerProGln 600  
1801 CAGGTAGCAGAGGCTATGATAAGAAGGTCGCGTGGTTTTCACATAGC 1850  
601 GlnValAlaGluAlaMetAspLysLysGlyAlaLeuValPheAspIleAl 617  
1851 TCTTATGTTCCCGTGGAGTTGTGTACGGTAACGGTGAAGTTTACTTGG 1900  
617 aLeuMetPheProValGluLeuLeuTyrGlyAsnGlyGluValTyrLeuG 634  
1901 AAGAAGCTCGATACGTTGGTCAAGAGGCAAGTCAATACCTGGCCTACAAT 1950  
634 luGluLeuAspThrLeuValLysArgGluGlyAspTyrLeuAlaTyrAsn 650  
1951 GTTGGTCAGTGTGGTGAAGTATGAACATTCCTTCTTAACGTAAGCGG 2000  
651 ValGlyGlnCysGlyGluMetTyrGluHisSerPheSerAsnValSerGI 667  
2001 GTTTTTCACCTTTTCTATGATAGCACTTCGTCGGGAACGTGTTAAGC 2050  
667 yPhePheThrPheSerTyrValArgThrSerSerGlyAsnValPheLysL 684  
2051 TAGAGTATAGGGATACCGTTGGTTGCTTACCATCATCTCACTATGCTGAGG 2100  
684 euGluTyrGluGlyTyrArgCysGlyTyrHisLeuThrMetCysArg 700  
2101 GCTCAGAAGTCACCTGGAACCTAGGTTTACGTATAGTTCGTTGGTCCGTC 2150  
701 AlaGlnLysSerProGlyThrGluValThrTyrArgSerLeuValProSe 717  
2151 GTTCGTGGGCAATTCGCTGTGTTACATCTGTTGTAGCTGGTTCAGTG 2200  
717 rPheValGlyLysSerLeuValPheIleProValValAlaGlySerSerV 734  
2201 TGTCTCTTAACACAAATAGTCCCTCGATTCGGACTTTGTCGACAGATCTAT 2250  
734 alSerPheLysThrIleValLeuAspSerAspPheValAspArgIleTyr 750  
2251 TCCTACGCGCTCAACACTATAGGACATTCGAGAATAGAAGCTTTAGTA 2300  
751 SerTyrAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPheGluTy 767  
2301 TGCCGTTGGGGGCTCAGGTGCGAAAGACCCATGTCATTACAGGGAGTC 2350  
767 rAlaValGlyAlaValArgSerGlnLysThrHisValIleThrGlySerA 784  
2351 GCCTTCTCCACACAGAGTTGATATTCCTCTGATGATATGTCGGGTTTA 2400  
784 rgValValHisSerLysValAspIleSerProAspAspMetTrpGlyLeu 800  
2401 GTTGTGCTCTTATGGCTCAGGCGATTAAAGGATAGGCGCAAGAGTATTCC 2450  
801 ValValAlaValMetAlaGlnAlaIleLysAspArgAlaLysSerIleAr 817  
2451 CTCTTATAACTTTTATAAAGCCAGTGAGGGAGTCTCGCCGGGCTCTTCA 2500  
817 gSerTyrAsnPheIleLysAlaSerGluGlySerLeuAlaGlyValPheL 834  
2501 AGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACCGAGTCTCGGTC

834 ysLeuPhePheGlnThrValGlyAspCysPheSerAsnAlaValSerVal 850  
2551 TATGCTAAGCAATGTCACGATTAACCTTCAACGTTTTTGGAGACGCTTAT 2600  
851 TyrAlaLysAlaMetValHisAspAsnPheAsnValLeuGluThrLeuMe 867  
2601 GTCTATCCCGAGACGCTTCATCCGTAAAGTACCTGGGCTCTGTTGTTA 2650  
867 tSerMetProArgAlaPheIleArgLysValProGlySerValValValT 884  
2651 CCATTTGCACCTTCGGAGCTTCAGACAGGTTGGAGCTCAGGGTGCCTTT 2700  
884 hrIleCysThrSerGlyAlaSerAspArgLeuGluLeuArgGlyAlaPhe 900  
2701 GATATTTCGAAGGAGACCTTCGGTAGGAACTGAAGATAGTCGCTTGGC 2750  
901 AspIleSerLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuAr 917  
2751 CGTCTTCTCTAGGCTATCGTGGAAATTCAAATTAAAGTCAATGAAGCAA 2800  
917 gValPheSerArgAlaIleValGluAspSerIleLysValMetLysAlaM 934  
2801 TGAAGACAGAAGATCGAAACCCCTGCCAATTACTGAAGATTCTCTATAT 2850  
934 etLysThrGluAspGlyLysProLeuProIleThrGluAspSerValTyr 950  
2851 CGCTTCAATAATGGGAACTTCTAACGTCACCTGTACGAGGCGCAGGTCT 2900  
951 AlaPheIleMetGlyAsnValSerAsnValHisCysThrArgAlaGlyLe 967  
2901 TCTTGGCGGTTTCGAAACGCGGTTTCGAGTGTTCCTAAGGGTTTGG 2950  
967 uLeuGlyGlySerLysAlaThrValValSerSerValSerLysGlyLeuV 984  
2951 TAGCTCGTGGGCTGCGACGAAGCCCTTTCTGCGATTACGTCCTTCTTT 3000  
984 alAlaArgGlyAlaIleThrLysAlaPheSerGlyIleThrSerPhePhe 1000  
3001 TCCACAGGTTCACTATTCTACGACCGCGGTTAACTGAAGATGAAGGCT 3050  
1001 SerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAspGluArgLe 1017  
3051 TGATGCTCTGGTGGCACAGAGATGCTATAAACTCACCGTGGGCATAC 3100  
1017 uAspAlaLeuValArgThrGluAsnAlaIleAsnSerProValGlyIleL 1034  
3101 TGGAGAGCTCGCGGTAGCTGTGAGCAAGGTCGTAGCTGGAACCAAGAA 3150  
1034 euGluThrSerArgValAlaValSerLysValValAlaGlyThrLysGlu 1050  
3151 TTTTGGAGTGAAGTTTCTTTAAATGACTTCACCACTTTTCGTTATTCGGA 3200  
1051 PheTrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAs 1067  
3201 TAAGGTGCTTATCGGGATATTCGTGGCTCTTTGGGTGCGGCCCAATTG 3250  
1067 nLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleA 1084  
3251 CATGGAAGTATAGCGCGGAATTCGGGCTAACCGCTAGAGGTACCGCGGC 3300  
1084 laTrpLysTyrArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGly 1100  
3301 AGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCGGTGTTT 3350  
1101 SerSerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLe 1117  
3351 AC CGGTTTAACTCTTAGCACAGTATCCGGTGGATCTTTAGTCGTGCCAA 3400  
1117 uArgGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgA 1134  
3401 GAGGGTTTTCGTGGCGGTGACCGTCACTAGGCGCACCGTAGCTAAACGT 3450

1134 rgGlyPheSerSerAlaValThrValThrArgAlaThrValAlaLysArg 1150  
3451 CAAAGTCCCCTTAGCGTTGCTATCGTTTCTTACTCATACGCCATTTCGGG 3500  
1151 GlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGI 1167  
3501 CTGCAGTATTAGGCAATTGGGCACATCTCTTCCACGGCACATTAAATGT 3550  
1167 yCysSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetP 1184  
3551 TTTTCTTTCGTTTAGGACATTGCTTGGCGGAGGGCTAGCGCGAATACT 3600  
1184 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThr 1200  
3601 TGGAACTTTGGAGGCTTCCCAATAATTGGTGGTGGTGGTGGTGGTGGT 3650  
1201 TrpLysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValVa 1217  
3651 TTGGCGAGGGAAGAGTGCAGCTCAATTGTTACTTGCCCTATTAGCGTAGGG 3700  
1217 LrpArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyV 1234  
3701 TATCTTTGATCATAGGGCTTGCTTAAGCACACCATACCTCAACTGGCT 3750  
1234 aISerLeuIleIleArgGlyLeuLeuLeuAsnAspThrIleProGlnLeuAla 1250  
3751 TAGCTCCCAACCGTAGAGGGAGGAATGTGTACGATGAGACGCTTAGGTA 3800  
1251 TyrValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTy 1267  
3801 TTACCGGGACTTTGACTATGACCAAGGTGCTGCTCCATCTGGGACATCAGC 3850  
1267 rTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnH 1284  
3851 ATGAACGGTTCGCCGTGACGATACGATGATGATGATGATGATGATGATG 3900  
1284 iSGluAlaValProGlyAspAspAsnAspGlySerThrSerSerValSer 1300  
3901 AGCTATGATGTTGTCAAAATGTGGCGACGTGGGATTAGCACCAACGG 3950  
1301 SerTyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGI 1317  
3951 GGAAGTTACTGGTGAAGAAGACACCATTCACCTCGAAGCGTCAATACA 4000  
1317 yGluValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrT 1334  
4001 CTATGTCAGGAAGAGGTGCGCGCTGCGAGCTGTCGCGGGAAGACAA 4050  
1334 hrTyrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGln 1350  
4051 GGTGATCCGTCGGGTCTGCTACCGCTGACGCTATGGCTTTTGTGAAG 4100  
1351 GlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSe 1367  
4101 TGTGAAAAAGGTGTCAGCATGCTTTTACCACACAGTCTAGTGGGAAA 4150  
1367 rValLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluT 1384  
4151 CGGCTCGTAGGTTAGGTGGAGCGCAAGGGTTGCTCCCAAGACGCTC 4200  
1384 hrAlaArgGluValGluValAspGlyLysGlyLeuLeuProGluSerVal 1400  
4201 GTCGCTGAGCGCCGCACACAAAGAGGGAAGCTGCAAGTGGTAACAC 4250  
1401 ValGlyGluAlaProThrGlnGluArgGlyArgAlaAlaAspGlyAsnTh 1417  
4251 AGCACAACCGCGGTCAACGAAGGCGACAGGAGCCAGCTACAGTCCAGTC 4300  
1417 rAlaGlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerL 1434  
4301 TTGTGAGTTCGCCAGAGCTGATATCCAAAGTCAACCGTCAACCGTAGGTA 4350  
1434 euValSerSerProGlnAlaAspIleProLysValThrGlnSerGluVal 1450

4351 CATGCTCAGAAAGAGTGAACAAGAAGTACCATTGGCGACTGTTTTCGGG 4400  
1451 HisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSerGI 1467  
4401 CCCCAGCCCAATCGTGGATGAGAAACCCGCCCAAGTGTACGACTCGTG 4450  
1467 yAlaThrProIleValAspGluLysProAlaProSerValThrThrArgG 1484  
4451 GTGTGAAGATAATTGACAAGGCGAAGCGCTCGCTCATGTGGCTGAGAAA 4500  
1484 lYValLysIleIleAspLysGlyLysAlaValAlaHisValAlaGluLys 1500  
4501 AAACAGGTACAAGTCGAGCAGGCCAACAGAGAGTTTGACGATCAATCA 4550  
1501 LysGlnValGlnValGluGlnProLysGlnArgSerLeuThrIleAsnGI 1517  
4551 AGCAAGCGCGGTAAACAGCTTTGTCATGTTAGAACGTGTTCTCGCGTG 4600  
1517 uGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSerCysGlyV 1534  
4601 TGCAGCTGGATGTGTACAAAGGACGACTATCGCCACAGGTTCTCAAAC 4650  
1534 aIGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsn 1550  
4651 GCATTTACCTTTGTCGATAAAGTGAAGGACGAGTGGCGTCTTTTCTC 4700  
1551 AlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPhePheSe 1567  
4701 AAAGTGGGTGAGGGGTATACCTATATATGTTAGCCATGTTTTCATCAG 4750  
1567 rLysLeuGlyGluGlyTyrThrTyrAsnGlySerHisValSerSerG 1584  
4751 GGTGGCTCGTGGCTAGAGGATATCTTAACGGCAATTAAAGTACCCAAAG 4800  
1584 lYTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSer 1600  
4801 GTCTTCGACCACTGTTAGTGCAGAGTACAAGATGGTGGAGCGCTACC 4850  
1601 ValPheAspHisCysLeuValGlnLysTyrLysMetGlyGlyValPr 1617  
4851 ATTCCACGCTGATGACGAGGAGTGTCTATCCATCAGATAACCCCTATCTGA 4900  
1617 oPheHisAlaAspAspGluCysTyrProSerAspAsnProIleLeuT 1634  
4901 CGGTCAATCTCGTGGGCAAGGCAAACTTCTCGACTAAGTGCAGGAAGGT 4950  
1634 hrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLysGly 1650  
4951 GGTAAAGTTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCC 5000  
1651 GlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMetPr 1667  
5001 TTGCGGTTTTCAAAGACGCACTTCGATTCAGTAAACTCCATCGACCAAG 5050  
1667 oCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspGluG 1684  
5051 GCGCATCAGTTTGACGTTACGGCAACTCGCGCGTCTTTGGTGTAGGC 5100  
1684 lYArgIleSerLeuThrPheArgAlaThrArgArgValPheGlyValGly 1700  
5101 AGGATGTTGACGTTAGCCGCGCGGTGTCGGATGAGAAGTCAACGAGTGT 5150  
1701 ArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSerProGlyVa 1717  
5151 TCCAAACAGCAACACAGAGCCAAAGGTGCTACCAAGAACAAATCACACAA 5200  
1717 lProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIleThrProL 1734  
5201 AATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGG 5250  
1734 ysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluValLysGly 1750

5251 AGGTCGACATCGATATATGTCGACAAAGATTACGTTAGGAAGTCTGA 5300  
|||||  
1751 ArgSerThrTyrSerIleIlePcysGluInAspTyrValArgLysCysGI 1767  
|||||  
5301 GTGGCTCAGGCGTGATAATCCAGTGTGCTTGAACCTGACTACACCC 5350  
|||||  
1767 uTTPLeuArgAlaAspAsnProValMetAlaLeuGluProAspTyrThr 1784  
|||||  
5351 CAATGACATTTGAAGTGGTTAAACCGGACCTCTGAAGATCGCGTGG 5400  
|||||  
1784 roMetThrPheGluValValLysThrGlyThrSerGluAspAlaValVal 1800  
|||||  
5401 GAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGCGTT 5450  
|||||  
1801 GluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLe 1817  
|||||  
5451 GCTTATGGCTAGAAATATTGCCGTCCTACCGCGGAGGTGTTCTGAAAG 5500  
|||||  
1817 uLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyValLeuLysV 1834  
|||||  
5501 TACCTAATCAAGTTTATGATCACTACCGGGCTTTCACGTTTACAGTCG 5550  
|||||  
1834 alProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSer 1850  
|||||  
5551 GGCACAGATCTCATTTTTCATTCAACACAGACGGCTTGCGTCTGACAGA 5600  
|||||  
1851 GlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAs 1867  
|||||  
5601 CCTACCGTACGTACTCATAGCTGAAAAGGTATCTTTACCAAGGGCAAG 5650  
|||||  
1867 pLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysA 1884  
|||||  
5651 ATGTCGACGGGTGGTAGCTTTGGGCGACAACTCTGTTGATCGGACGAT 5700  
|||||  
1884 spValAspAlaValAlaLeuGlyAspAsnLeuPheValCysAspAsp 1900  
|||||  
5701 ATACTGGTTTCCACCATGCCATTAATTTTCATAGCTGCCAGTGAAGTCG 5750  
|||||  
1901 IleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAl 1917  
|||||  
5751 TCGATCGGCGATGGTGGGGAATCGTTTAAAGTCCTTCGAATATAAGTCG 5800  
|||||  
1917 aArgCysGlyMetValGlyGluSerPheLysSerPheGluTyrLysCyst 1934  
|||||  
5801 ATAATGCTCCCGCAGGTGGGTAAGACGACGACGCTAGTGGACGAATTC 5850  
|||||  
1934 yrAsnAlaProProGlyGlyGlyLysThrThrThrLeuValAspGluPhe 1950  
|||||  
5851 GTTAACTCACCAATAGCACAGCCACCATACGCTAATGTGGGAAGTTC 5900  
|||||  
1951 ValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySerSe 1967  
|||||  
5901 TGAGGACATAAATATGCGGTGAAGAGAGATCCGAATTTGGAAGGTC 5950  
|||||  
1967 rGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGlyL 1984  
|||||  
5951 TCAACAGTGTACCAAGTTAACTCAGGGTGGTAAACTTTATCGTCAGG 6000  
|||||  
1984 euAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg 2000  
|||||  
6001 GGAATGTATAAAGGTTTGGTGGATGAGGTCCACATGATGATCAAGG 6050  
|||||  
2001 GlyMetTyrLysArgValLeuValAspGluValIleHisMetMetHisGlnI 2017  
|||||  
6051 CTTACTACAACTAGGCGCTTTCGCAACCGCGCGCTCGGAAGGCTCTTTT 6100  
|||||  
2017 yLeuLeuGlnLeuGlyValIlePheAlaThrGlyAlaSerGluGlyLeuPheP 2034  
|||||  
6101 TTGGAGACATAAATCAGATACCATTTTCATAAACAGGAGAGGAGGTGTTAGG 6150  
|||||  
2034 heGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArg 2050  
|||||  
6151 ATGGATTGCTGCTTTTGTTCCAAGAAGGAAAGCGTTGTATACACTTC 6200  
|||||

2051 MetAspCysAlaValPheValProLysLysGluSerValValTyrThrSe 2067  
|||||  
6201 TAAATCGTACAGGTGTCGGTTAGATGTTTGTCTACTTGTCTCTCAATGA 6250  
|||||  
2067 rLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetT 2084  
|||||  
6251 CCGTAGGGGAACGGAAAGTGTACCTGAAAAGTCGTTAGCGGTAAG 6300  
|||||  
2084 hrValArgGlyThrGluLysCysTyrProGluLysValValSerGlyLys 2100  
|||||  
6301 GACAAACCACTAGTAAAGTCGTCCTCAAAAGGCCAATTCGAACCACTGA 6350  
|||||  
2101 AspLysProValValArgSerLeuSerLysArgProIleGlyThrThrAs 2117  
|||||  
6351 TGACGTAGCTGAATAAACGCTGACGCTGTACTTGTGCATGACCCAGTTGG 6400  
|||||  
2117 pAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuG 2134  
|||||  
6401 AGAAGTCGGATATCAAGAGGTCTGTTGAAGGAAAGAAAGAAACACCA 6450  
|||||  
2134 lLysSerAspMetLysArgSerLeuLysGlyLysGlyLysGluThrPro 2150  
|||||  
6451 GTGATGACACTGCATGAAGCACAGGGAACAAACATTCAGTCATGTGTTATT 6500  
|||||  
2151 ValMetThrValHisGluAlaGlnGlyLysThrPheSerAspValValLe 2167  
|||||  
6501 GTTTAGGACGAAGAACGCCGATGACTCCCTATTCACTAAACAAACCCGATA 6550  
|||||  
2167 uPheArgThrLysLysAlaAspAspSerLeuPheThrLysGlnProHisI 2184  
|||||  
6551 TACTTGTGTTTGTGCGAGACACAGCTGCTGCTGTTTATGCGGCTCTG 6600  
|||||  
2184 leLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeu 2200  
|||||  
6601 AGCTCAAGTTGGACGATAAGGTGCGGCACATATATTAGCCACGCGTCACC 6650  
|||||  
2201 SerSerLysLeuAspAspLysValGlyThrTyrIleSerAspAlaSerPr 2217  
|||||  
6651 TCAATCAGTATCGAGCTTTGCTTTCACAGCTTCGCCCGCGCTGGTTGCT 6700  
|||||  
2217 oGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysP 2234  
|||||  
6701 TTCGAGGTATA 6711  
|||||  
2234 heArgGlyIle 2237  
|||||

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT:W21636

seq\_documentation\_block:  
ID W21636 standard; Protein; 1390 AA.  
XX  
AC W21636;  
XX  
DT 19-SEP-1997 (first entry)  
XX  
DE Grapevine leafroll virus helicase.  
XX  
KW GLRaV; grapevine; Vitis; rootstock; leafroll; disease resistance;  
KW transgenic plant; tristeza virus; citrus; helicas.  
XX  
OS Grapevine leafroll associated virus type 3 isolate NY1.  
XX  
PN W09722700-A2.  
XX  
PD 26-JUN-1997.  
XX  
PF 20-DEC-1996; 96WO-US20747.  
XX  
PR 21-DEC-1995; 95US-0009008.  
XX  
PA (CORR ) CORNELL RES FOUND INC.  
XX

PI Gonsalves D, Ling K;  
 XX WPI; 1997-341691/31.  
 DR N-PSDB; T72214.  
 XX  
 PT DNA encoding grape-vine leaf-roll virus proteins - useful to impart  
 PT viral-resistance to Vitis scion or root-stock cultivar(s)  
 XX  
 PS Claim 4; Page 69-75; 172pp; English.  
 XX  
 CC A helicase (W21636) was identified as the expression product of  
 CC open reading frame la (T72215) of a double-stranded cDNA molecule  
 CC derived from grapevine leafroll associated virus type 3 (GLRAV-3)  
 CC isolate NY-1 dsRNA. The helicase, and other GLRAV-3 polypeptides  
 CC (see also W21637-47), can be expressed in transgenic plants as a  
 CC means of imparting GLRAV resistance to Vitis scion or rootstock  
 CC cultivars or tristeza virus resistance to citrus scion or rootstock  
 CC cultivars, or used to raise antibodies useful for detection of  
 CC GLRAV.  
 XX  
 SQ Sequence 1390 AA;

alignment\_scores:  
 Quality: 5247.00 Length: 1423  
 Ratio: 4.214 Gaps: 17  
 Percent Similarity: 87.491 Percent Identity: 74.139

alignment\_block:  
 US-09-301-906-3 x W21636 ..

Align seg 1/1 to: W21636 from: 1 to: 1390

```

2542 GNCCTCGCTATGCTAAGCAATGTGCAGCATAACTCAACGTTTGGCA 2591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 ValSerThrTyrAlaLysSerValMetAsnAspAsnPhcAsnIleuG1 17

2592 GACGCTTATGCTATGCCAGAGCGCTTCATCGTAAAGTACCTGGGTCG 2641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 uThrLeuValThrLeuProLysSerPheIleValLysValProGlySerV 34

2642 TTGCTGTTACCAATTCGACTTCTGGAGCTTCAGACAGGTGGAGCTCAGG 2691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 aLeuValSerIleThrThrSerGlyIleSerAspLysLeuLeuLeuArg 50

2692 GGTGCTTTGATATTGCAAGGAGACCTTCGGTAGGAACTCAAGAATAG 2741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GlyAlaPheAspValSerLysLysAsnPheSerArgArgLeuArgSerSe 67

2742 TCGCTTGGCGCTTCTCTAGGCGCTATCGTGAAGATTCAATTAAGGTCA 2791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 rArgLeuArgValPheSerArgAlaIleValGluAspThrIleLysValM 84

2792 TGAAGCAATGAACAGACAGATGGAACACCCCTGCCAATTAAGAT 2841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 etLysGlyMetLysSerGluAspGlyLysProLeuProIleLeuAlaGluAsp 100

2842 TCTGTATATGCTTCATATGGGAGCGTTCTTAACGTCCACTGTACGAG 2891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 SerValTyrAlaPheMetThrGlyAsnMetSerAsnValHisCysThrAr 117

2892 GGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTCGAGTGTTCCTA 2941
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 gAlaGlyLeuLeuGlyGlySerLysAlaCysAlaLeuSerLeuAlaValL 134

2942 AGGGTTTGTAGTCTGGGGCTGGCAGACGAGCGCTTTCTGGCATACG 2991
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 ysGlyAlaAlaSerArgAlaThrGlyThrLysLeuPheSerGlyLeuThr 150

2992 TCGTTCTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACCTGAAGA 3041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 SerPheLeuSerAlaGlyLcylLeuPheThrAspGluGlyLeuThrProGl 167

```

```

3042 TGAAGGCTTGATGCTCTGGTCGCACAGAGAAATGCTATAAACTCACCGG 3091
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 yGluArgLeuAspAlaLeuThrArgArgGluHisAlaValAsnSerProv 184

3092 TGGGCATACCTAGAGAGCTCGCGTAGCTGTGAGCAAGTCGTAGCTGGA 3141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 aIGlyLeuLeuGluProGlyAlaSerValAlaLysArgValValSerGly 200

3142 AGGAAAGAAATTTTGGAGTGAAGTTCTTAAATGACTTCACCACCTTTCCT 3191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 ThrLysAlaPheLeuSerGluLeuSerLeuGluAspPheThrThrPheVa 217

3192 ATTGGCGAATAAGTGTCTTATCGGATATTCTGGCGCTCTTTGGGTCCG 3241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 lIleLysAsnArgValLeuIleGlyValPheThrLeuSerMetAlaLeuT 234

3242 CCCCAATTGCTGGAAGTATAGCGCGGAATTGCGGCTAACGCTAGACGG 3291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 hrProValValTrpLysTyrArgArgAsnIleAlaArgThrGlyValAsp 250

3292 TACGCGGCGAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAGCGCG 3341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 ValPheHisArgAlaArgSerGlyThrAlaAlaIleGlyLeuGlnCysLe 267

3342 CGGTGGTTTACCGGTTTAACTCTAGCACAGTATCCGCTGGATCTTTAG 3391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 uSerGlyGlyArgSerLeuAlaGlyAspAlaAlaArgGlyAlaLeuThrV 284

3392 TCGTGGGAAGAGGGTTTTCGTCGGGGTACCGTCACTAGGGCGACCGTA 3441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 aThrArgGlyGlyLeuSerSerAlaValAlaValThrArgAsnThrVal 300

3442 GCTAAAGTCAAGTCCCTTAGCTGTGCTATCGTTTCTACCTCATACGC 3491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 AlaArgArgGlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAl 317

3492 CATTTCCGGCTCGAGTATGTTAGGCATTTGGGCACATGCTTTCACCGCG 3541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
317 aValSerGlyCysThrLeuLeuGlyIleTrpAlaHisAlaLeuProArgH 334

3542 ACTTAATGTTTTCGTTTGGTTTAGGACATTTGCTGGGCGGAGGCGTAGC 3591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
334 lSerMetPhePheGlyLeuGlyThrLeuPheGlyValSerAlaSer 350

3592 GCGAATCTTGAAGTTTGCAGGCTTCTCCAATAATTTGGTCGCTCTCC 3641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 ThrAsnSerTrpSerLeuGlyTyrThrAsnSerLeuPheThrValPr 367

3642 CGAGTTCTTTCGCGGAGGAGAGTGTACGCTCATTTTACTGCCTATTTA 3691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
367 oGluLeuThrTrpGluGlyArgSerTyrArgSerLeuLeuProGlnAlaA 384

3692 CCCTAGGGGTATCTTTGATCATAGGGGCTTCTTAACGACACCATACCT 3741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
384 laLeuGlyIleSerLeuValValArgGlyLeuLeuSerGluThrValPro 400

3742 CAACCTGTCTACGTCGCCACCGTAGAGGAGAAATGTGTACGATGAGAC 3791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 GlnLeuThrTyrValProProlleGluGlyArgAsnValTyrAspGlnAl 417

3792 GCTTAGGTATTACCGGACTTTTCACTATGACGAAGTGTGTGCTCCATC 3841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
417 aLeuAsnPheTyrArgAspPheAspTyrAspGlyAlaGlyProSerG 434

3842 GGACTCAGCATGAAGGGTTCGCGGTGAGCATACGATGGATCCACTTCT 3891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
434 lYThrAlaGlyGlnSerAspProGly...ThrAsnThrSerAspThrSer 449

3892 AGTGTCTCAAGCTATGATGTTCTCACAAATGTGCGGACGTGGGATTAG 3941
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
450 SerValPheSer.....AspAspGlyLeuPr 458

3942 CACCAACGGG.....GAAGTTACTGGTGAAG 3967

```

.....  
458 oAlaSerGlyGlyGlyPheAspAlaArgValGluAlaGlyProSerHisA 475  
3968 AAGAGACCCATTACCTCGA...AGCGTGAATACACTTATGTCGAGGAA 4014  
:::||||| |||||:::||||| |||||  
475 laValAspGluSerProArgGlySerValGluPheValTyrArgGluArg 491  
4015 ...GAGTTGCCCGCTGCTGAGCTGGCGGAAAGACAGAGTAT... 4056  
::: |||||::: |||||::: |||||::: |||||  
492 ValAspGluHisProAlaCysGlyGluAlaGluValGluLysAspLeuII 508  
4057 ...CCGTCGGTCTGCTACCGCTGACGCT...A 4084  
||| |||||::: |||||::: |||||::: |||||  
508 eThrProLeuGlyThrAlaValLeuGluSerProProValGlyProGluA 525  
4085 TGCGTTTGTGAAGTGTGAAAAAGGTGTCGACGATGCTTTTACCAA 4134  
::: |||||::: |||||::: |||||::: |||||  
525 laGlySerAlaProAsnValGluAspGlyCysProGluVal... 538  
4135 CAGCTAGTGGGNAACGGCTCGTGAAGTTGAGTGGACGGCAAGGTT 4184  
::: |||||::: |||||::: |||||::: |||||  
539 .....GluAlaGluLysCysSerGluValIleValAsp... 549  
4185 GCTCCAGAAAGCGTCTGCGTGGCGCACAGAAGAGGGAAGAG 4234  
::: |||||::: |||||::: |||||::: |||||  
550 .ValProSerSer...GluProValGlnGluValLeuGluS 563  
4235 CTCAGATGGTAAACAGACACAAACCGGCTCAACGAAGCGACAGGAG 4284  
::: |||||::: |||||::: |||||::: |||||  
563 erThrAsnGly...ValGlnAlaAlaArgThrGlu...Glu 574  
4285 CCAGTACAGTCCAGTCTTGTGATGTCGCCACAGCTGATATTCGAAGGT 4334  
|||||::: |||||::: |||||::: |||||  
575 ValValGlnGlyAspThrCysGlyAla...GlyVa 585  
4335 CACCCAGTCGAGGTACATCTCAGAAAGAGTGAAACAGAGAGTACCA 4383  
|:::||||| |||||::: |||||::: |||||  
585 lAlaLysSerGluVal...SerGlnArgValPheProAlaGlnValProA 601  
4384 .....TTGGCGACTGTTTCGGGCGCC... 4404  
||| |||||::: |||||::: |||||::: |||||  
601 laHisGluAlaGlyLeuGluAlaSerSerGlyAlaValValGluProLeu 617  
4405 .....ACGCCAATCGTGAATGAGAACCCGCCCAAGTGT...AC 4442  
::: |||||::: |||||::: |||||::: |||||  
618 GlnValSerValProValAlaValGluLysThrValLeuSerValGluLy 634  
4443 GACTGTGGTGAAGATAATTGACAAGGCGCAAGCGCTCGCTCATGTGG 4492  
::: |||||::: |||||::: |||||::: |||||  
634 sAlaArgGluLeuLysAlaValAspLysGlyLysAlaValAlaHisAlaL 651  
4493 CTGAGAAAAAAGGTACAGTCCAGCAGCCCAACAGAGAGTTCGACG 4542  
||| |||||::: |||||::: |||||::: |||||  
651 ysGluValLysAsnValProValLysThrLeuProArgGlyAlaLeuLys 667  
4543 ATCAATGAGGCAAGCGCGTGAACAGCTTGTGATGTTAGACGTGTC 4592  
|||||::: |||||::: |||||::: |||||  
668 IleSerGluAspThrValArgLysGluLeuLysMetPheArgThrCysSe 684  
4593 CTGCGGTGTCAGCTGATGTCACAGCAAGCGACTATCGCCACAGAGT 4642  
|||||::: |||||::: |||||::: |||||  
684 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 701  
4643 TCTCAACGCATTACCTTTGTCGATAACTTGAAGGGAGGTTCGGTC 4692  
|||||::: |||||::: |||||::: |||||  
701 heSerAsnAlaPheThrPheValAspSerLeuLysGlyArgSerAlaVal 717  
4693 TTTTCTCAAAGCTGGGTAGGGGTATACCTATAATGGTGGTACCATGT 4742  
|||||::: |||||::: |||||::: |||||  
718 PhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisVa 734  
4743 TTCATCAGGTGCGCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGT 4792  
|||||::: |||||::: |||||::: |||||

734 lSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysT 751  
4793 ACCCAAGCGTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGTGA 4842  
|||||::: |||||::: |||||::: |||||  
751 yrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 767  
4843 GGGTACCATTCCACGCTGATGACGAGAGTGCTATCCATCAGATAACCC 4892  
|||||::: |||||::: |||||::: |||||  
768 GlyValProPheHisAlaAspAspGluGluCysTyrProSerAspAsnPr 784  
4893 TATCTTACGCTCAATCTCTGGGGAAGGCAAACTCTCGACTAAGTGCA 4942  
|||||::: |||||::: |||||::: |||||  
784 oIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysA 801  
4943 GGAAGGTGCTAAGTCATGCTATAAACGCTAGCTTCGGGTGACTATTTT 4992  
|||||::: |||||::: |||||::: |||||  
801 rglYsGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 817  
4993 CTTATGCTTGGGTTTCAAAGGACGCACTTGCAATTCAGTAAACTCCAT 5042  
|||||::: |||||::: |||||::: |||||  
818 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerII 834  
5043 CGACGAAGGGCGCATCAGTTTGAGTTTCAGGGCAACTCGGCGCTCTTG 5092  
|||||::: |||||::: |||||::: |||||  
834 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArgValPheG 851  
5093 GTGTAGCAGGATGTTGCAGTTAGCCGGCGGTGTCGGATGAGAAGTCA 5142  
|||||::: |||||::: |||||::: |||||  
851 lyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSer 867  
5143 CGAGGTGTTCCAAACAGCAACACAGAGCCAGGTGCTACCAGAACAT 5192  
|||||::: |||||::: |||||::: |||||  
868 ProGlyValProAsnGlnProGlnSerGlnGlyAlaThrArgThrII 884  
5193 CACACAAAATCGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGCAAG 5242  
|||||::: |||||::: |||||::: |||||  
884 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 901  
5243 TCAAGGGGAGTGCACATACCTCGATATGTCGCAACAAAGATTACGTTAGG 5292  
|||||::: |||||::: |||||::: |||||  
901 alLysGlyArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArg 917  
5293 AGTGTGAGTGGCTCAGGGCTGATAATCCAGTGTATGCTCTTGAACCTGA 5342  
|||||::: |||||::: |||||::: |||||  
918 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuLysProGI 934  
5343 CTACACCCCAATGACATTTCAAGTGTTAAACCGGCACTCTCAAGATG 5392  
|||||::: |||||::: |||||::: |||||  
934 yTyrThrProMetThrPheGluValValLysAlaGlyThrSerGluAspA 951  
5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTTGAGAGGACATAC 5442  
|||||::: |||||::: |||||::: |||||  
951 laValValGluTyrLeuLysTyrLeuAlaIleGlyIleGlyArgThrTyr 967  
5443 AGGGCTTGTCTTATGGCTAGAAATATTGGCTGCTCACTACCGCGGAGGTGT 5492  
|||||::: |||||::: |||||::: |||||  
968 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyVa 984  
5493 TCTGAAAGTACCTAACTCAAGTTTATGAATCACTACCGGCTTTCCACGTTT 5542  
|||||::: |||||::: |||||::: |||||  
984 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1001  
5543 ACAAGTCGGGCACAGATCTCATTTTTCATTTTCAACACAAAGACGGCTTCGCT 5592  
|||||::: |||||::: |||||::: |||||  
1001 yrLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArg 1017  
5593 GTGAGAGACCTACCGTACGTACTCATAGCTGAAAGGATATCTTTACCAA 5642  
|||||::: |||||::: |||||::: |||||  
1018 ValArgAspLeuProTyrValPheIleAlaGluLysGlyIlePheIleLy 1034  
5643 GCGCAAGATGTCGACCGGTGGTGGCTTTGGCGCAACTCTGTTCTGAT 5692  
|||||::: |||||::: |||||::: |||||  
1034 sGlyLysAspValAspAlaValValAlaLeuGlyAspAsnLeuSerValC 1051

```
5693 GGCACGATATCTGGTTTCCACGATGCCATTAAATTTGATAGTGCACATG 5742
|||||
1051 ysAspAspIleLeuValPheHisAspAlaIleAsnLeuMetGlyAlaLeu 1067
|||||
5743 AAAGTCCTCGATCGCGCATGTGGCGGAATCGTTTAACTCCTTCGAATA 5792
|||||
1068 LysValAlaArgCysGlyMetValGlyGlySerPheLysSerPheGluTyr 1084
|||||
5793 TAAGTCTATAATGCTCCCGAGGTGGCGTAAAGACGACGCTTAGTCGG 5842
|||||
1084 rLysCysTyrAsnAlaProProGlyGlyGlyThrThrMetLeuValAla 1101
|||||
5843 AGCAATTCTTAAGTCACCCATAGCACAGCCACCATTAACGGCTAATGTG 5892
|||||
1101 spGluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1117
|||||
5893 GGAAGTTCGAGACATAAATATGCGGTGAACAGAGAGATCCGAATTT 5942
|||||
1118 GlySerSerGluAspIleAsnMetAlaValLysLysArgAspProAsnLe 1134
|||||
5943 GGANGCTCTCAACAGTCTACACAGTAACTCCAGGCTGTAACCTTTA 5992
|||||
1134 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAlaAsnPheI 1151
|||||
5993 TGCTCAGGGAATGTATAAAGGGTTTTCGTGCATCAGCTGCACATGATG 6042
|||||
1151 leValArgGlyMetTyrLysArgValLeuValAspLysValTyrMetMet 1167
|||||
6043 CATCAAGGCTTACTACAACCTAGCGTCTCCGAACCGCGCTCGGAAGG 6092
|||||
1168 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluG 1184
|||||
6093 CCTCTTTTGGAGACATAAATCAGATACCATTCATAAACAGGAGAGG 6142
|||||
1184 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 1201
|||||
6143 TGTGTTAGGATGATGCTGCTGTTTGTTCCTCAAGAAAGGCGGTTGTA 6192
|||||
1201 aPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 1217
|||||
6193 TACACTTCTAAATCGTACAGGTGTCGGTTAGATGTTTGTCTACTGTTGTC 6242
|||||
1218 TyrThrLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 1234
|||||
6243 CTCATGACCGTAAGGGAACGGAAGGTTTACCTGAAAGGTCGTTA 6292
|||||
1234 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 1251
|||||
6293 GCGGTAAAGGACAAACAGTAAAGATCGCTGCTCCAAAAGGCAATTTGA 6342
|||||
1251 erGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 1267
|||||
6343 ACACGTGATCAGCTAGCTGAATAAACGCTGACGTGCTACTTGTGATGAC 6392
|||||
1268 ThrThrAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThr 1284
|||||
6393 CCAGTTGGAGAAGTCGGATATCAAGAGGTCGTTGAAGGGAAGGAAGAAG 6442
|||||
1284 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLSG 1301
|||||
6443 AAACACCACTGATGACAGTGCATAGACACAGCGGAAACAAATTCAGTGAT 6492
|||||
1301 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 1317
|||||
6493 GTGCTATTCTTTAGCAGGAAGAAGCCGATGACTCCCTATTTCACATAACA 6542
|||||
1318 ValValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysG 1334
|||||
6543 ACCGATATACCTGTTGGTTTTCGAGACACACAGCTCAGCTGTTTANG 6592
|||||
1334 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 1351
|||||
```

```
6593 CCGCTCTAGCTCAAAAGTTGGACGATAAGTTCGGCAGACATATATTAGCGAC 6642
|||||
1351 laAlaLeuSerSerGluLeuAspAspLysValGlyThrTyrIleSerAsp 1367
|||||
6643 GGTTCACCTCAATCAGTATCCGACGCTTTCGTTTCACACGTTGCCCCGGC 6692
|||||
1368 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 1384
|||||
6693 TGGTTGCTTTCGAGGTATA 6711
|||||
1384 aGlyCysPheArgGlyIle 1390
|||||

seq_name: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT:Y58145
seq_documentation_block:
ID Y58145 standard; Protein; 291 AA.
XX
AC Y58145;
XX
XX 07-MAR-2000 (first entry)
XX
DE GLRaV-3 polyprotein helicase domain.
XX
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody.
XX
OS Grapevine leafroll-associated virus 3.
XX
PN WO9955880-A1.
XX
PD 04-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09307.
XX
PR 29-APR-1998; 98US-0083404.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Ling K;
XX
DR WPI: 2000-062035/05.
DR N-PSDB; 249205.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 15; Fig 11; 84pp; English.
XX
CC This sequence represents the helicase domain of the
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotide
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
```

CC GLrav-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.  
XX

SQ Sequence 291 AA;

alignment\_scores:

Quality: 1495.00 Length: 291  
Ratio: 5.137 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-301-906-3 x Y58145

Align seg 1/1 to: Y58145 from: 1 to: 291

```
5764 GTGGCGCAATCGTTTAAAGCTTCGCAATATAAGTCGTATAATGCTCCCCC 5813
|||||
1 ValGlyGluSerPheLysSerPheGluTyrLysCysTyrAsnAlaProP 17
|||||
5814 AGGTGGCGGTAGACGACGACGCTTACTGGACGAATTCGTTAAGTCACCCA 5863
|||||
17 oGlyGlyLysThrThrThrLeuValAspGluPheValLysSerProA 34
|||||
5864 ATAGCACACCCACCATACGGCTAATCTGGAGAGTTCTGAGGCATAAAT 5913
|||||
34 snSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsn 50
|||||
5914 ATGGCGGTGAAGACAGAGATCGAATTTGGAAGGCTCAACAGTCTAC 5963
|||||
51 MetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSerIleAla 67
|||||
5964 CACAGTTAACTCCAGGGTGTAACTTTATCGTCAGGGGAATGTATPAAAA 6013
|||||
67 rThrValAsnSerArgValValAsnPheIleValArgGlyMetTyrLysA 84
|||||
6014 GGGTTTTGGTGGATGAGTGCACATGATGCATCAAGCTTACTACAACTA 6063
|||||
84 rgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeu 100
|||||
6064 GCGCTCTCCCAACCGCGCGTCCGGAAGCGCTCTTTTGGGACATAAA 6113
|||||
101 GlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAs 117
|||||
6114 TCAGATACCATTCATAAAGGAGAGAGTGTAGGATGGATTGTGCTG 6163
|||||
117 nGlnIleProPheIleAsnArgGluLysValPheArgMetAspCysAlaV 134
|||||
6164 TTTTGTTCCTCAAGAGAAAGCGTTGTATACACTTCTAAATCGTACAGG 6213
|||||
134 aPheValProLysLysGluSerValValTyrThrSerLysSerTyrArg 150
|||||
6214 TGTCCGTTAGATCTTGTCTACTTGTCTCCTCAATCACCGTAAGGGGAAC 6263
|||||
151 CysProLeuAspValCysTyrLeuLeuSerSerMetThrValArgGlyTh 167
|||||
6264 GGAAGAAGTGTACCTCGAAAGGCTGTTAGCGGTAAAGCAACACAGTAG 6313
|||||
167 rGluLysCysTyrProGluLysValSerGlyLysAspLysProValV 184
|||||
6314 TTAGATCGCTGCCAAAGGCCAATTTGGAACACACTGATGACGTAGCTGAA 6363
|||||
184 aArgSerLeuSerLysArgProIleGlyThrThrAspAspValAlaGlu 200
|||||
6364 ATAAAGCTTGACGTGTACTGTTCATGACCCAGTTCGAGAAGTCGATAT 6413
|||||
201 IleAsnAlaAspValIrrLeucysMetThrGlnLeuGluLysSerAspMe 217
|||||
6414 GAAGAGCTCGTTCAAGGGAAAAAGAAACACACAGTGTGACAGTGC 6463
|||||
217 tLysArgSerLeuLysGlyLysGlyLysGluThrProValMetThrValH 234
|||||
6464 ATGAAGCACAGGAAAAACATTCAGTGTGTGTATGTTTAGGACGAAG 6513
```

```
|||||
234 isGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLys 250
|||||
6514 AAAGCCGATGACTCCCTATTTCACCTAAACACCGCATATACTTGTGGTTT 6563
|||||
251 LysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGlyLe 267
|||||
6564 GTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAAGTTGG 6613
|||||
267 uSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuA 284
|||||
6614 ACCATAAGGTCGCGACATATATT 6636
|||||
284 sPAspLysValGlyThrTyrIle 291
```

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:Y58144

seq\_documentation\_block:

ID Y58144 standard; Protein: 272 AA.

XX Y58144;

XX 07-MAR-2000 (first entry)

XX GLrav-3 polyprotein methyltransferase domain.

XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLrav-3;  
KW viral disease; yield loss; sugar content; inhibition; infection;  
KW replication; polyprotein; domain; proteinase; methyltransferase;  
KW helicase; RNA-dependent; RNA polymerase; untranslated region;  
KW transgenic plant; component; resistant; truncation; deletion; antisense;  
KW expression; detection; antibody.

XX OS Grapevine leafroll-associated virus 3.

XX PN WO9955880-A1.

XX PD 04-NOV-1999.

XX PF 29-APR-1999; 99WO-US09307.

XX PR 29-APR-1998; 98US-0083404.

XX PA (CORR ) CORNELL RES FOUND INC.

XX PI Gonsalves D, Ling K;

XX DR WPI; 2000-062035/05.

XX DR N-PSDB; 249204.

XX PT Newly isolated grapevine leafroll virus protein or polypeptide useful  
XX for producing transgenic plants conferring viral disease resistance -

XX PS Claim 5; Fig 8; 84pp; English.

XX CC This sequence represents the methyltransferase domain of the  
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3  
CC (GLrav-3). Leafroll is a serious viral disease, occurring wherever  
CC grapes are grown. Although not lethal, it causes yield losses and  
CC reduction in sugar content. The virus encodes several proteins,  
CC which may serve as targets for the inhibition of viral infection or  
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),  
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase  
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase  
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF  
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.  
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',  
CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are  
CC useful for the generation of transgenic plants and plant components.  
CC Such transgenic plants may be resistant to viral disease, for example,  
CC this property being conferred on the plants via the use of nucleotides  
CC encoding truncated or internally deleted proteins, or via the use of  
CC antisense nucleotides to inhibit viral gene expression. The nucleotides



CC may additionally be used in the detection of viral nucleic acids in a  
CC tissue sample. The proteins can be used to generate antibodies which  
CC may be used to detect GLRAV-3 in plant samples. The isolation of  
CC GLRAV-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.  
XX

SQ Sequence 272 AA;

alignment\_scores:  
Quality: 1432.00 Length: 272  
Ratio: 5.265 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-301-906-3 x Y58144 ..

Align seq 1/1 to: Y58144 from: 1 to: 272

1378 CTGAAGCCGCGGGAAGGAGAGAGCTGAGGGAACCTCTTTCCAGAGCTTTC 1427  
1 LeuLysProArgGluArgGluLysLeuArgGluLeuPheProGluLeuSe 17  
1428 GATACAGTTCTCCGACTCGGTCAGAGTAGTCACCCATTCGCTAATGCCA 1477  
17 rIeGlnPheSerAspSerValArgSerSerHisProPheAlaAsnAlaM 34  
1478 TSCGAGGCTGTTCAATGGAATCTTTTCCAGAGGCTGCTAATCTGTC 1527  
34 eLArgSerCysPheAsnGlyIlePheSerArgCysGlyAsnValCys 50  
1528 TTCTTCGATATTGGGGGAGGCTTCACGATCATGTCAAAGCTGGCCATGT 1577  
51 PhePheAspIleGlyGlySerPheThrTyrHisValLysAlaGlyHisVa 67  
1578 GAACGTGTCATGATGCATCCAGTCCTAGACGCTTAAAGATGTGAAGCGGA 1627  
67 lAsnCysHisValCysAsnProValLeuAspValLysAspValLysArgA 84  
1628 CAATCAATCAGATCTCTCTTTCTTCCACAGCTGGGGAGATTCGTAACGTG 1677  
84 rGleAsnGluLeuLeuPheLeuSerThrAlaGlyGlyAspSerTyrVal 100  
1678 TCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGCTTACTGTAG 1727  
101 SerSerAspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCysSe 117  
1728 TCGAAGTCGAGAACTCGGATCTAGACCGCATCCGGGTTTATGGTGG 1777  
117 rArgGluSerGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValA 134  
1778 ATGTCTACGATATATCCCGCAGCAGGTAGCAGAGGCTATGCATAAGAAG 1827  
134 spValTyAspIleSerProGlnGlnValAlaGluAlaMetAspLysLys 150  
1828 GGTGCGCTGGTTTTCCAGATAGCTCTTATGTTCCCGCTGGAGTTGTTGTA 1877  
151 GlyAlaLeuValPheAspIleAlaLeuMetPheProValGluLeuLeuTy 167  
1878 CGGTAACGGTGAAGTTTACTTGGGAAGACTCGATACGTTGGTGGTAAGAGGG 1927  
167 rGlyAsnGlyGluValTyLeuGluGluLeuAspThrLeuValLysArg 184  
1928 AAGGTGATTACCTGCGCTACAAATGTTGCTCAGTGGTGAGATGATGAA 1977  
184 luGlyAspTyrLeuAlaTyrAsnValGlyGlnCysGlyGluMetTyrGlu 200  
1978 CATTCCTTCTCTAAGTGAAGCGGGTTTTTCACCTTTTCTTATGTACGCAC 2027  
201 HisSerPheSerAsnValSerGlyPhePheThrPheSerTyrValArgTh 217  
2028 TTCGTCGGGGAACGTGTTTAACTAGCTAGATATGAGGGATACCGTTGCTGTT 2077

217 rSerSerGlyAsnValPheLysLeuGluTyrGluGlyTyrArgCysGlyT 234  
2078 ACCATCATCTCACTATGTGTAGGCTCAGAAGTCACTGGAAGTGAAGTT 2127  
234 yrHisLeuThrMetCysArgAlaGlnLysSerProGlyThrGluVal 250  
2128 ACGTATAGTCTGTTGCCCTCGTTCGTGGGCAAAATCGCTGCTGTTCA 2177  
251 ThrTyrArgSerLeuValProSerPheValGlyLysSerLeuValPheIl 267  
2178 ACCTGTTGTAGCTGGT 2193  
267 eProValValAlaGly 272

seq\_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W73476

seq\_documentation\_block:

ID\_W73476 standard; Protein; 2639 AA.

XX W73476;

XX AC

XX 29-MAR-1999 (first entry)

XX DE

XX KW

XX KW

XX OS

XX PN

XX PD

XX PF

XX PR

XX PA

XX PI

XX DR

XX DR

XX PT

XX PS

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

SQ Sequence 2639 AA;  
  
 alignment\_scores:  
     Quality: 802.50 Length: 2475  
     Ratio: 0.681 Gaps: 115  
 Percent Similarity: 47.596 Percent Identity: 22.424  
  
 alignment\_block:  
 US-09-301-906-3 x W73476 ..  
  
 Align seg 1/1 to: w73476 from: 1 to: 2639  
  

```

454 CAAGATGACGACAAGCCCAACCGGAAGCGTCACATTGGCGCTGTAAGCC 503
|||||: ||| |||||: :||| :||| :|||:
324 GlnSerValIstYrLysArgGluAspAlaHisArgThrValGluGl 340
504 AACAGCTGTTGGCGTCCAGCTACCCTCCTCTAAAAAACAGAACACTGG 553
|||||: ||| ||| |||||: :||| :||| :|||:
340 uargAlaLaIGlyGlySerValGluGlnProArgGlnLysArgIleaspG 357
554 AGCCA...CGCAATCAGTCCCACACAGCTCGTTG..... 585
|| :|||: |||||: :||| :||| :|||:
357 LuLysGlyCysGlyArgValProSerGlyGlyPheSerHisLeuLeuVal 373
586 .....GAGGAGAAGCGCGCTTGACGCTTTGGCCCTTTCTTCAG 623
||| :||| :||| :||| :||| :|||:
374 GlyAsnLeuAasnGluValArgArgLysValAlaAlaGlyLeuLeuArgPh 390
624 TAAAGTGGCGGTGATGAGAGCGCACGCTGTCATCTTCGCGAAAGGAAAT 673
||| :|||: |||||: :||| :||| :|||:
390 eargValGlyGlyAspMetasp..... 397
674 TGTTTTAACAGGCCCTTAATGTTCTCTATTGATGATAAGAACAGCTTCGTT 723
|||||: |||||: :||| :||| :||| :|||:
398 ..PheHisArgSerPheSerThrGlnAlacly..HisArgLeuLeuVal 412
724 TGGGCTAAA.....ATCTGGGATGAAGCCTC 749
||| :||| :|||: |||||: :||| :|||:
413 TrpArgArgSerArgSerValCysLeuGluLeuTyrrSerProSerLy 429
750 TCCTAGGAGAGCGGTATTTTACGTCAA.....GATAGAGCTCTTA 790
:||| :||| :||| :||| :|||:
429 sasnPheLeuArgTyrrAspValLeuProCysSerGlyAspTyrrAlaAlam 446
791 AATTCTTCCTATTGTGCGGGTAGGCTACGATCGAGGACTTCATCGTG 840
||| :||| :|||: |||||: :||| :|||:
446 etPheSerPheAlaAlaGlyGlyArgPheProLeu..... 457
841 AATACAGCCCCAGGTGTGATGTGCCTTCGCCCGCATTTGAGTTGTGGAG 890
||| :||| :|||: |||||: :||| :|||:
458 .....ValLeuMetThrArgIleArgTyrr.... 465
891 TATGCGCGAAAGGCGCTTTGTATGCACCACCAAGGCTGCTGTGGTTTA 940
||| :||| :|||: |||||: :||| :|||:
466 .....ProAsnGlyPheCystyrLeuA 473
941 ACAATGACAGGCTGAGGGGAGAAATTTACAGACGCTGCTGCTCTCACT 990
:||| :||| :||| :||| :||| :|||:
473 IaHisCysArgtyrAlaLacysAlaPheLeuLeuArgGlyPheAspProLys 489
991 TCCTTTTCGATAGGT.....TT 1007
|||||: |||||: :||| :||| :||| :|||:
490 ArgPheAspIleGlyAlaPheProThrAlaAlaLysLeuArgAsnArgMe 506
1008 CTTGATGACACCTTGGCTTTTAGATCGTTAAAGGTCATTAGTWTTCGGGCA 1057
:||| :|||: |||||: :||| :||| :|||:
506 tValSerGluLeuGlyGluArgSerLeu..... 515
1058 CGAACATACTACATGCCATCCTCACTCAATGACAGCGTACCTTTGGGTGG 1107
||| :||| :|||: |||||: :||| :|||:
516 .....GlyLeuAsnLeuTyrrGlyAlaTyrrThrSer 525
```

[illegible]

808 SerSerCysTyrSerHisLysLeuSerIleIleLysAspIleMetThrTh 824  
2013 TTCTTATGTACCACTCTCCGCAACGTGTTTAAAGCTAGATGATGAGG 2062  
824 rProTyrLeu...ThrLeuGlyGlyPheLeuPheSerValGluMetTyrG 840  
2063 GATACCGTTGTGTACCATCATCTCACTATGCTAGGGCTCAGAAGTCA 2112  
840 LuValArgMetGlyValAsnTyrPheLysIleThrLysSerGluValSer 856  
2113 CTGGAAGCTAGGTTACG...TATAGTCTGTTG... 2142  
857 ProSerIleSerCysThrLysLeuLeuArgTyrArgArgAlaAsnSerAs 873  
2143 .....GTCCCGCTGCTGCTGGGCAAAATCGCTGGTGTCA 2176  
873 pValValLysValLysLeuProArgPheAspLysLysArgMetCysL 890  
2177 TACCTGTGTAGCTGTTCTAGTGTGCTCTTTAAGACAAATAGCTCGAT 2226  
890 euPro.....GlyTyrAspThrIleTyrLeuAsp 899  
2227 TCGGACTTTGTCACAGATCTATTCTACGCGCTCAACACTATAGGGAC 2276  
900 SerLysPheValSerArgValPheAspTyrValValCysAsnCysSerAl 916  
2277 ATTCCAGAAATACAGCTTTTGACTATGCCGTGGCGGTCAGTCGCAAA 2326  
916 aValAsnSerLysThrPheGluTrpValTrpSerPheIleLysSerSerL 933  
2327 AGACCCATGTCTATTACAGGAGTCGCGTTGTCCACAGCAAGTTGATATT 2376  
933 ysSerArgValIleIleSerGlyLysIleIleHisLysAspValAlaLeu 949  
2377 TCTCTGTATGATATGCGGTTTACTTGTGCTGTATGCTGATAGGCGCAT 2426  
950 AspLeuLysTyrValIleSerPheAlaAlaValMetLeuAlaSerGlyVa 966  
2427 TAAGGATAGG.....GCGAAGAGTATTCCCTCCTATA 2458  
966 lArgSerArgLeuAlaSerGluTyrLeuAlaLysAsnLeuSerHisPheS 983  
2459 ACTTTATAAAGCCAGTGAGGGAGTCTCGCGGGGTCTTCAAGCTCTTC 2508  
983 er..... 983  
2509 TTTCAGCGGTAGGCGATTGT...TTTTCGACGAGCTCTCGTCTATGC 2555  
984 .....GlyAspCysSerPheIleGluAlaThrSerPheValLe 996  
2556 TAAGGCAATGGTGACGAT.....AACTTCAACGTTTGGAGACGC 2596  
996 uArgGluLysIleArgAsnMetThrLeuAsnPheAsn.....GluArgL 1011  
2597 TTATGCTATGCCAGAG.....GGTTTCATCCGTAAGTACCTGGTCT 2640  
1011 euLeuGlnLeuValLysArgValAlaPheAlaThrLeuAspValSerPhe 1027  
2641 GTTGTGTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGCGCTCAG 2690  
1028 LeuAspLeuAspSerThrLeuGluSerIleThrAspPheAlaGluCysLy 1044  
2691 GGGTGCCTTGTATTTCCGAAG.....GAGACCTTCG 2722  
1044 sValAlaIleGluLeuAspGluLeuGlyCysLeuArgAlaGluAlaGluA 1061  
2723 GTAGGAACCTGAAGAATAGTCGCTTTCGCGGTCTTCTAGGCTATCGTG 2772  
1061 snGluLysIleArgAsn.....LeuAla 1068  
2773 GAAGATTCAATTAAAGTCATGAAGCAATGAAGACAGAA.....GA 2813  
||||||| :||| :|||

1069 GlyAspSerIleAlaAlaLysLeuAlaSerGluIleValValAspIleAs 1085  
2814 TGGAAAAACCCCTCCCAATTACTGAAGATTCTGTATATGCGTTTCATATGG 2863  
1085 pSerLysProSerProLysGln.....ValG 1094  
2864 GGAACGTTTCTTAACGTCCACTGT.....ACGAGGGCAGGT 2898  
1094 lyAsnSerSerSerGluAsnAlaAspLysArgGluValGlnArgProGly 1110  
2899 CTTCTTGGCGGTTTCGAAACCGACCGGTGTTTCGAGTGTCTTAAGGGTTT 2948  
1111 LeuArgGlyGlySerArgAsnGlyValValGlyGlu..... 1122  
2949 GGTAGTCTGCTGGGCTGCGACGAAGCCCTTTCTGCGCATTAAGTCTGCTCT 2998  
1123 .....PheLeuHisPheV 1127  
2999 TTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAG 3048  
1127 alValAspSerAlaLeuArgLeuPheLysTyrAlaThrAspGlnGlnArg 1143  
3049 CTTGATGCTCTGTCGCCACACAGAAATGCTATAAATCAACCGTGGGCAT 3098  
1144 IleLysSerTyrValArg.....PheLeuAspSerAlaValSerPh 1157  
3099 ACTGGAGACGTCGCGGTAGCTGTGAGCAAGTCTGTAGCTGGAACAAAG 3148  
1157 euAspTyrAsnTyrAspAsnLeuSerPheIleLeuArgValLeuSerG 1174  
3149 AATTTTGGAGT.....GAAGTTTCTTAAATGA 3176  
1174 LuGlyTyrSerCysMetPheAlaPheLeuAlaAsnArgGlyAspLeuSer 1190  
3160 ..... 3176  
1191 SerArgValArgSerAlaValCysAlaValLysGluValAlaThrSerCy 1207  
3177 CTTCCACCACTTTGCTATTTCGGAATAAGTGTCTATCCTGGGATATTCTGG 3226  
1207 salAsnAlaSerValSerLysAlaLysValMetIle..... 1219  
3227 CGTCTTTGGTGGCGCCCAATTTGCATGGAAGTATAGGCGC...GGAATT 3273  
1220 ..ThrPheAlaAlaAlaValCysAlaMetMetPheAsnSerCysGlyPhe 1235  
3274 GCGGCTAACCTAGAAAGTACGCGGC..... 3300  
1236 SerGlyAspGlyArgGluTyrLysSerTyrIleHisArgTyrThrGlnVa 1252  
3301 .....AGTAGTTACGAACTCTAAGCT 3322  
1252 lleuPheAspThrIlePhePheGluAspSerSerTyrLeuProIleGluV 1269  
3323 CGTTAAGTTTCAACAGCCCGGTGTTTACGCGGTTTAACTCTTACCTCTAGC... 3369  
1269 alLeuSerSerAlaIleCysGlyAlaIleValIleThrLeuPheSerSerGly 1285  
3370 .....ACAGTATCCGCTGATCTTTAGTCGTCGGAAGAGGGTTTTC 3410  
1286 SerSerIleSerLeuAsnAlaPheLeuLeuGlnIleThrLysGlyPheSe 1302  
3411 GTCGCGGTGACCGTCACTAGGCGCACCGCTAGCTAAACGCTCAAGTCCCT 3460  
1302 rLeuGluValValValArgAsnValVal.....ArgValThrH 1315  
3461 TAGCGTTGCTATCGTTTCTACCTCATAGCCCATTTCCGGCTG..... 3503  
1315 isGlyLeuSerThrThrAlaThrAspGlyValIleArgGlyValPheSer 1331  
3504 CAGTATGTTAGGCATTG.....GGCAGTCTCTTCCACGCA..... 3542  
1332 GlnIleValSerHisLeuLeuValGlyAsnThrGlyAsnValAlaTyrGl 1348  
||||| :||| :|||

```

3543 .....CTTA 3546
1348 nSerAlaPheIleAlaGlyValValProLeuLeuValLysLysCysVals 1365
3547 ATGTTTTCCTTTAGGACATGCTTGGGCGAGGCTAGCGGAA 3596
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1365 erLeuIlePheIleLeuArgGluAspThrTyrSerGlyPheIleLysHis 1381
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3597 TACTTGAAGTTGGAGCTTCCCAATAATGTCGCTGTTCCCGAGG 3646
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1382 GlyIleSerGluPheSerPheLeuSerSerIleLeuLysPheLeuLysGI 1398
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3647 TTGTTTGGCAGGAGAGTGTGAGTATGTTTACTGCTATACGTA 3696
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1398 yLysLeuValAspGluLeuLysSerIleLeuLysPheLeuLysGI 1409
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3697 GGGTATCTTTGATCATAGGGGCTTGAACGACACCATACCTCAACT 3746
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1409 InGlyValPheAspSerAsnLysHisValPheLysGluAlaThrGlnGlu 1425
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3747 TGCTTACGTCACCGGTA..... 3765
1426 AlaIleArgThrThr.ValMetGlnValProValAlaValAlaAspAlaL 1442
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3766 .....GAGGAGGAATGTGTACGATGAGAGCGTTCAGTATTAC 3804
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1442 eLysSerAlaAlaGlyLysIleTyrAsnAsnPheThrSerArg..... 1456
3805 CGGACTTTGACTATGACGAAGTGTGTCATCTGGGACTCAGCATGA 3854
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1457 ArgThrPheGlyLysAspGluGlyProGlyLeuLysGlyLysSers 1473
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3855 AGCGGTTCCGGTGACGATAACGATGA...TCCACTTTAGTGTCTCAA 3901
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1473 uTyrPheSerCysAspGluGlyProGlyLeuLysGlyLysSers 1490
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3902 GCTATGATGTTGCACAAATGTGC.....CGGACGTGGGGATT 3939
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1490 erTyrGly.PheSerIleLeuAlaPhePheSerArgIleMetTrpGly.. 1505
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3940 AGCACAACGGGAGTACTGTGAGGAGAGACACCATTCACCTCGAAG 3989
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1506 .....AlaArgArgLeuIleValLys.....ValLysHisGluC 1517
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
3990 CGTCAATACACTTATGTCGAGGAGAGGTTTGCC.....CGT 4027
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1517 yspPheGlyLysLeuPheGluPheLeuSerLeuLysLeuHisGluPheArg 1533
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4028 CTCGACGTGGCGGAAGACAAAGGTGATCCGTCGGGTTCTGTGACCGCT 4077
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1534 ThrArgValPhe.GlyLysAsnArgThrAspValGlyValTyrAspPheL 1550
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4078 GACGCTATGCTTCTGCTGAAAGTGTCAAAAAGGTTCCACGATGCTTT 4127
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1550 euProThrGly.IleValGluThrLeu..... 1558
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4128 TCACCAACAGTCTAGTGGGGAACAGGCTCGTGAGTTGAGTGGACGGCA 4177
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1559 .....SerIleGluGluCysAspGlnIleGlu..... 1568
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4178 AAGGTTGCTCCAGAAAGCGTGTGAGGAGCGCGCACACAAAGAAAG 4227
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1569 ..GluLeuLeuGlyAspAspLeuLysGly..... 1577
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4228 GGAAGAGTGCACATGCTAACACAGACACAAACCGCGTCAACGAAGCGA 4277
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1578 .....AspLysAspAlaSerLeuThrAspMetAsnTyrPheGI 1590
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4278 C...AGGAGCCAGTACAGTCCAGTCTGTGAGTTCGCCACAGGCTGATA 4324
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1590 uPheSerGluAspPheLeuAlaSerIleGluGluProProPheAlaGlyL 1607
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382

```

```

4325 TTCAAAGGTCACCCAGTCC..... 4344
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1607 euArgGlySerLysAsnIleAlaIleLeuAlaIleLeuGluTyrAla 1623
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4345 .....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1624 HisAsnLeuPheArgIleValAlaSerLysCysSerLysArgProLeuPh 1640
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4383 ATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCC 4432
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1640 eLeuAlaPheAlaGluLeuSerSerAlaLeuIleGluLys.....PheL 1655
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4433 CAAGTGTTCAGCTGCTGCTGAGATATTTGACAAGGCGAAGCCGCTC 4482
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1655 ysGluValPheProArg.....LysSerGlnLeuVal 1665
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4483 GCTCATGTGCTGAGAAACACAGGTACAGTCGAGCAGCCCAACACAGAG 4532
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1666 AlaIleValArgGluTyrThrGln...ArgPheLeuArgSerArgMetAr 1681
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4533 GAGTTTGACGATCAATGAAGCAAGCGCGTAAACAGCTTTGCATGTTTA 4582
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1681 gAlaLeuGlyLeuAsnAsnGluPheVal.ValLysSerPheAlaAspLeu 1697
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4583 GAACGTGTTCTCGGGTG.....TGCAGCTGGATGTGTACAAC 4620
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1698 .....LeuProAlaLeuMetLysArgLysValSer.GlySerPheLeuA 1712
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4621 GAAGCGACTATCGCCACGAGTCTCAACGCGATTTTACCTTTGTCGATAA 4670
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1712 laserValTyrArg...ProLeuArgGlyPheSerTyrMetCysValSer 1727
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4671 CTTGAAGAGGAGGAGTGGCGTCTTTCTCAAGCTGGGTGAGGGGTATA 4720
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1728 AlaGluArgArgGluLysPhePheAlaLeuValCysLeuIleGlyLeuSe 1744
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4721 CCTATATGCTGTGAGCCATGTTTCATCAGG..... 4751
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1744 rLeu.....ProPhePheValArgIleValGlyAlaLysAlaC 1757
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4752 .....GTGGCTCGTCCCTAGAGATATCTTAACGGCAATTAA 4790
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1757 ysGluGluLeuValSerSerAlaArgArgPheTyr..... 1768
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4791 GTACCAAGCGTCTTCA..... 4808
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1769 .....GluArgIleLysIlePheLeuArgGlnLysTyrValSerLeuSe 1783
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4809 .....CCACTGTTTGTGACAGAGTACAAGAT..... 4835
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1783 rAsnPhePheCysHisLeuPheSerSerAspValAspSerSerAlas 1800
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4835 ..... 4835
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1800 erAlaGlyLeuLysGlyAlaSerArgMetThrLeuPheHisLeuLeu 1816
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4836 .....GGGTGAGGCGGTACCATTC 4855
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1817 ValArgLeuAlaSerAlaLeuLeuSerLeuGlyTrpGluGlyLeu..... 1831
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4856 ACCTGATGACGAGGAGTGTCTCATTCATCAACCCCTATC...TTGACG 4902
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1832 .....LysLeuLeuLeuSer.HisHisAsnLeuLeuPheLeuCys 1844
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4903 GTCAATCTCGTGGGAGGCAAACTTCTCAGTAAGTCAGGAGGAGGTTG 4952
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1845 PheAlaLeuValAspAspValAsnValLeuIleLysValLeuGlyGlyLe 1861
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4953 TAAGGTGATGCTCATAAACGTAGTCTCG..... 4980
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
1861 uSerPheValGlnProIlePheSerLeuPheAlaMetLeuLeuG 1878
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382
4981 .....GGTGTACTAT..... 4989
.....GAGTACATGCTCAGAAAGAGTGAACAAAGAGTACC 4382

```

1878 InProAspArgPheValGluTyrSerGluLysLeuValThrAlaPheGlu 1894  
:::|::|  
4990 TTTCCTATGCGCTGC.....GGTTTCA 5012  
|::|::|::|::|::|::|::|::|::|::|::|::|  
1895 PhePheLeuLysCysSerProArgAlaProAlaLeuLeuLysGlyPhe.. 1910  
|::|::|::|::|::|::|::|::|::|::|::|::|  
5013 AAGGACCGCACTTCCATTCACTAACTCCATCGACGAAAGGCCGATCAGTT 5062  
:::|::|::|::|::|::|::|::|::|::|::|::|  
1911 .....PheGluCysValAlaAsnSerThrValSerL 1921  
5063 TCAGCTTCAGGCCAACTCGCGCGCTCTTT.....GGT 5094  
|::|::|::|::|::|::|::|::|::|::|::|::|  
1921 ysThrValArgArgLeuLeuArgCysPheValLysMetLeuLysLeuArg 1937  
5095 GTAGCAGCATGCTTCAGTTAGCCGCGC...GGCGTTCGGATGAGAAGTC 5141  
|::|::|::|::|::|::|::|::|::|::|::|::|  
1938 LysGlyArgGlyLeuArgAlaAspGlyArgGlyLeuHisArgGlnLysAl 1954  
|::|::|::|::|::|::|::|::|::|::|::|::|  
5142 A...CCAGGTGTTCCAAACACAGCAACACAGCAAGGTGCTACAGAA 5188  
:|::|::|::|::|::|::|::|::|::|::|::|  
1954 aValProValIleProSerAsnArgValValThrAspGlyValGluArgL 1971  
5189 CAATCACACCAAAATCGGGGGCCAGGCTCTATCT..... 5223  
:::|::|::|::|::|::|::|::|::|::|::|::|  
1971 euSerValLysMetGlnGlyValGluAlaLeuArgThrGluLeuArgIle 1987  
5224 .....GAGGGAAGTGTAGGNAAGTCAA 5246  
1988 LeuGluAspLeuAspSerAlaValIleGluLysLeuAsnArg....Ar 2003  
5247 GCGGAGGTTCGACATCTCGATATGTGCGCAACAAGATTACGTTAGGAAG. 5295  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2003 gAsnArgAspThrAsn.....AspAspGluPheThrArgProA 2016  
5296 .....TGTGAGTGGCTCAGGGCT 5313  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2016 laHisGluGlnMetGlnGluValThrThrPheCysSer.....LysAla 2030  
5314 GATAATCCAGTATGGCTCTTGAA..... 5337  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2031 AsnSerAlaGlyLeuAlaLeuLeuArgAlaValLeuValGluAspAlaI 2047  
5338 .....CCTGACTACACCCCAATGACATTTGAAGTGGTTAAACCGGG. 5379  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2047 eLysSerGluLysLeuSerLysThrValAsnGluMetValArgLysGlyS 2064  
5380 .....ACCTCTCAAGATGCCGTCGTG.....CAGTACTTG 5409  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2064 exThrThrSerGluValAlaValAlaLeuSerAspGluAlaVal 2080  
5410 AAGTATCTGGCTATAGGCATTGAGAGG.....ACATACAG 5444  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2081 GluGluIleSerValAlaAspGluArgAspSerProLysThrValAr 2097  
5445 G.....CGCTTGTCTTATGCTAGAA 5464  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2097 gIleSerGluTyrLeuAsnArgLeuAsnSerSerPheGluPheProLysP 2114  
5465 ATATTTCGCGTCACTACCGCCGGAAGGTTCGTGAAGTACCTAATCAAGTT 5514  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2114 roIleValValAspAspAsnLysAspThrGlyLeuThrAsnAlaVal 2130  
5515 TATGAATCACTACCGGGCTTTCACGTTTACAGTCGGCAGCATCTCAT 5564  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2131 ArgGlu.....PheTyrTyrMetGlnGluLeuAlaLeuPheGl 2143  
5565 TTTTTCATTCAACA.....CAAGACGGCTTGGCTGTG..... 5595  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2143 uIleHisSerLysLeuCysThrTyrTyrAspGlnLeuArgIleValAsnP 2160  
5596 .....AGAGACCTACCGTACCTGCATCA 5619  
:::|::|::|::|::|::|::|::|::|::|::|::|

2160 heAspArgSerValAlaProCysSerGluAspAlaGlnLeuTyrValArg 2176  
5620 GCTGAAAAGGTATCTTTACCAAGGCCAAGATGTCGACGCG...GTGGT 5666  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2177 LysAsnGlySerThrIleValGlnGlyLysGluValArgLeuHisIleLy 2193  
5667 AGCTTTGGCGGACAACTCTGTTGATGCGACGATATACTG..... 5706  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2193 sAspPheHisAspHisAspPheLeuPheAspGlyLysIleSerIleAsnL 2210  
5707 .....GTTTTCACGATGCCATTAAATTTGATA 5733  
2210 ysArgArgArgGlyGlyAsnValLeuTyrHisAsp.....AsnLeu... 2223  
5734 GTGCACTGAAAGTCGCTCGATCGCGCATGGTCGGCGAATCGTTT..... 5778  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2224 ...AlaPheLeuAlaSerAsnLeuPheLeuAlaGlyTyrProPheSerAr 2239  
5779 .....AAGTCTCTTCGAATATAAGTGTCTATAATGCTC 5809  
2239 gSerPheValPheThrAsnSerSerValAspIleLeuLeuTyrGluAlaP 2256  
5810 CCCCAGGTGGCGGTAAAGACGACAGCTTAGTGGACGAATTCGTT..... 5853  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2256 roProGlyGlyGlyLysThrThrLeuIleAspSerPheLeuLysVal 2272  
5854 .....AAGTCAACCAATAGCACAGCCACCATTACGGCTAATCTCGGAAG 5897  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2273 PheLysLysGlyGluValSerThrMetIleLeuThrAlaAsnLysSerSe 2289  
5898 TTCTCAGGACATAAATATATGCGGTGAAGAAGAGAGATCCGAATTTTGAA. 5946  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2289 rGlnValGluIleLeuLysLysValGluLysGluValSerAsnIleGluC 2306  
5947 .....GGTCTCAACAGTGTCTACACAGTT 5970  
2306 ysGlnLysArgLysAspLysArgSerProLysLysSerIleTyrThrIle 2322  
5971 AACTCCAGGTGGTAACTTTATCGTCAGGGGAATGTATAAAAGGTTTT 6020  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2323 AspAlaTyrLeuMetHisHis.....ArgGlyCysAspAlaAspValle 2337  
6021 G...GTGATAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCG 6067  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2337 uPheIleAspGluCysPheMetValHisAlaGlySerValLeuAlaCysI 2354  
6068 TCTTCGAACACCGCGCGTCGGAAGGCTCTTTTGGAGACATAATCAG 6117  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2354 leGluPheThrArgCysHisLysValMetIlePheGlyAspSerArgGln 2370  
6118 ATACCATTCATAACAGG.....GAGAAGGTGTTAGGATGGATTG 6158  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2371 IleHisTyrIleGluArgAsnGluLeuAspLysCysLeuTyrGlyAspLe 2387  
6159 TGCTGTTTTTTTGTCCAAAGAAAGCGTTGTATACACTCTTAATTCGT 6208  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2387 uAspArgPheVal...AspLeuGlnCysArgValTyrGlyAsnIleSerT 2403  
6209 ACAGTGTCTCGTTAGATGTTGCTACTTGTGTCTCTCAATG..... 6249  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2403 yrArgCysProIrpAspValCysAlaIrpLeuSerThrValTyrGlyAsn 2419  
6250 .....ACCGTAAGGGGAACGGNAAGTGTACCCTGAAAAGGTCGT 6290  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2420 LeuIleAlaThrValLysGlyGluSerGlu..... 2429  
6291 TAGCGGTAGGAACAAACCACTAGTAGATCGCTGTCCAAAAGGCCAATTG 6340  
|::|::|::|::|::|::|::|::|::|::|::|::|  
2430 ....GlyLysSer.....SerMetArgIleAsnGluIleA 2440  
6341 GAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTGTGTCATG 6390  
:::|::|::|::|::|::|::|::|::|::|::|::|  
2440 snSerValAspAspLeuValProAspValGlySerThrPheLeuCysMet 2456

```
6391 ACCAGTTGGAGAGTCGGATATGAAGAGTCGTTGAAGGGAAGA... 6438
    ||| ||||| : : : : : : : : : : : : : : : : : :
2457 LeuGlnSerGluLysLeuGluIleSerLysHisPheIleArgLysGlyLe 2473
6439 .AAGAAACACCACTGATGACAGTCATGAAGACACAGGGAACAACTTCA 6487
    : : : : : ||||| : : : : : : : : : : : : : : : :
2473 uThrLysLeuAsnValLeuThrValHisGluAlaGlnGlyGluThrTyra 2490
6488 GTGATGTGGTATTGTTAGACCAAGAAAGCCGATGACTCCCTATTCACT 6537
    : : ||| ||| ||| ||| : : : : : : : : : : : :
2490 laArgValAsnLeuValArgLeuLysPheGlnGluAspGluProPheLys 2506
6538 AAACACCCGATATCTGTTGGTTCGAGACACACAGCTCACTGGT 6587
    : : ||||| ||||| : : : : : : : : : : : : : : : :
2507 SerIleArgHisIleThrValAlaLeuSerArgHisThrAspSerLeuTh 2523
6588 TTATGCGCTCTGAGCTCAAGTTGGACGATAAGTTCGCGCACATATATTA 6637
    : : ||| : : : : : ||| : : : : : : : : : : : :
2523 rTyAsnValLeuAlaAlaArgArgGlyAspAlaThrCysAspAlaIleG 2540
6638 GCGACCGCTCA 6648
    : : |||||
2540 lnLysAlaAla 2543
```

seq\_name: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:Y91969

```
seq_documentation_block:
ID Y91969 standard; Protein; 309 AA.
XX
AC
XX
XX
XX
DT 19-JUL-2000 (first entry)
XX
DE PMWav-2 helicase.
XX
KW Pineapple mealybug wilt virus 2; PMWav-2; ORF 1a; helicase;
KW transgenic pineapple; resistance; antiviral.
XX
OS Pineapple mealybug wilt virus 2.
XX
PN WO200017372-A2.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22152.
XX
PR 23-SEP-1998; 98US-0101461.
XX
PA (UYHA-) UNIV HAWAII.
PA (UYFL) UNIV FLORIDA.
XX
PI Hu JS, Karasev AV, Dawson WO, Melzer M;
XX
DR WPI: 2000-283596/24.
DR N-PSDB: A08690.
XX
PT Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT useful for protecting pineapples against the virus
XX
PS Claim 17; Page 97-98; 112pp; English.
XX
CC Pineapple mealybug wilt virus 2 (PMWav-2), open reading frame (ORF) 1a
CC is incomplete but encodes this protein which encompasses all motifs
CC characteristic of viral helicases. The DNA and protein sequences
CC are useful for production of transgenic pineapple plant cultivars and
CC also in transformation methods to impart resistance against the virus to
CC pineapple plants. Mealybug wilt is a major problem limiting profitable
CC pineapple production in many pineapple growing areas worldwide, the
CC present invention aims to overcome this problem.
XX
SQ Sequence 309 AA;
```

```
alignment_scores:
    Quality: 631.50      Length: 304
    Ratio: 2.782        Gaps: 4
    Percent Similarity: 74.671      Percent Identity: 44.737
alignment_block:
US-09-301-906-3 x Y91969 ..
Align seg 1/1 to: Y91969 from: 1 to: 309
5845 GAATTCGTTAAGTCACCAATAGACAGCCACCATTACGGCTATGTGGG 5894
    |||:|:| |||:|:| : : : : : : : : : : : : : : |
6 GlutyrPheLysAspArgSerGlySerCysIleMetThrAlaAsnArgGI 22
5895 AAGTCTCTGAGCACATAATATGCGGTGAAGACAGAGATCCG..... 5937
    |||:|:| |||:|:| : : : : : : : : : : : : : : |
22 ySerAlaIleAspIleAsnAspThrIleGluSerIleAspAlaAlaAsnA 39
5938 .....AATTGGAGGTCCTCAACAGT..... 5958
    |||:|:| |||:|:| : : : : : : : : : : : : : : |
39 laSerLysAlaAlaSerAsnAsnValSerGlyValGluSerIleAspAsn 55
5959 .....GCTACCACAGTTAACTCCAGGTCGTAACTTTATCGTCAG 5999
    ||| ||| ||||| ||||| : : : : : : : : : : : :
56 TyrValCysAlaArgThrValAsnSerGlnIleMetAsn.....CysLy 70
6000 GGGAAATC.....TATAAAAGGGTTTGGTCGATGAGTGCACATGATGC 6043
    :|||:|:| ||| : : : ||||| ||||| : : : : : : |
70 sGlyValMetAsnTyrThrCysAlaLeuValAspGluMetTyrLeuMeth 87
6044 ATCAAGGCTTACTACAACTAGGCGTTCGCAACCGCGCGTCGGAAGGC 6093
    |||:|:| ||||| ||||| : : : : : : : : : : : :
87 lsLysGlyLeuMetLeuGlyValPheSerSerGlyAlaArgAla 103
6094 CTCTTTTTCGAGACATAATACATACATTCATATAACAGGAGGAGGT 6143
    :|||:|:| ||||| ||||| : : : ||||| ||||| : : : |
104 IlePheTyrGlyAspIleAsnGlnIleProPheIleAsnArgGluLysCy 120
6144 GTTTAGGATGGATTGCTGCTTTTGTTCCAAAGAGGAAGCGTTCAT 6193
    ||| : : : |||:|:| ||| ||| ||||| ||||| : : : |
120 sPheTyrSerLysGluGlyValTyrCysProGlyLysAspGluIleIle 137
6194 ACATTTCTAAATCGTACAGTGCCTGTAGATGTTGCTACTTGTGTCTCC 6243
    |||||:|:| ||||| ||||| : : : ||||| ||||| : : : |
137 yrThrSerGluSerTyrArgCysProAlaAspValCysMetTyrSerSer 153
6244 TCAATGACCGTAAGGGGACGGAAGAGTGTACCTGAAAGGTCGTTAG 6293
    |||:|:| : : : : : : : : : : : : : : ||| : : : |
154 SerLeuLysAlaGlnAlaGlySerAsnArgTyrLeuLysGlyValSerCy 170
6294 CGGTAAGGACAAACCACTAGTAAGTCGCTCCAAAGGCCAATTGGAA 6343
    : : : : : : : : : : : ||||| ||||| ||||| ||||| : : : |
170 sAsnGlnArgGluValValLeuArgSerLeuSerLysArgProValValt 187
6344 CCACTGATGACGTAGCTGAATAAACGCTGACGTGTTGTCATGACACC 6393
    |||||:|:| ||| : : : : : ||||| ||||| : : : : : |
187 yrAlaGluGlnIleGlnLeuGluAlaAspAlaTyrIleThrPheLys 203
6394 CAGTTGGAGAAAGTCGGATATGAAGAGGTCGTTGANGGGAAGGAAGA 6443
    ||| |||:|:| : : : : : ||||| ||||| : : : : : |
204 GlnGluCysLysGluLysValValArgAlaLeuArgAlaValGlyArgAr 220
6444 AACACCACTGATCATGAGTCATGAAGCACAGGGAACAAACATTTCAGTGC 6493
    : : : : : ||||| ||||| ||||| ||||| : : : : : |
220 gasPlysValPheThrSerHisGluAlaGlnGlyMetThrPheGlyArgV 237
6494 TGGTATTGTTAGGACGAAGAAAGCCGATGACTCCCTATTTCATAACAA 6543
    ||||| ||| ||| : : : ||||| ||||| : : : : : : : |
237 alValLeuCysArgLeuSerAlaThrAspSerValPheSerSerGlu 253
6544 CCGCATATACTTGTGTTGTCGACACACACAGGTCACCTGGTTATGC 6593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```





XX  
PA (UYHA-) UNIV HAWAII.  
PA (UYFL) UNIV FLORIDA.  
XX  
PI Hu JS, Karasev AV, Dawson WO, Melzer M;  
XX  
XX WPI; 2000-283596/24.  
DR N-PSDB; A08684.  
XX  
XX Isolated pineapple mealybug wilt virus proteins and polypeptides,  
PT useful for protecting pineapples against the virus  
XX  
PS Claim 5; Page 79-82; 112pp; English.  
XX  
XX Pineapple mealybug wilt virus 1 (PWVav-1), open reading frame (ORF) 1a  
CC encodes this protein with all 8 motifs conserved in the so-called viral  
CC helicases. The DNA sequences and related proteins are useful for  
CC production of transgenic pineapple plant cultivars and also in  
CC transformation methods to impart resistance against the virus to  
CC pineapple plants. Mealybug wilt is a major problem limiting profitable  
CC pineapple production in many pineapple growing areas worldwide, the  
XX present invention aims to overcome this problem.  
SQ Sequence 598 AA;

## alignment\_scores:

Quality: 323.00 Length: 490  
Ratio: 1.192 Gaps: 22  
Percent Similarity: 55.306 Percent Identity: 26.735

## alignment\_block:

US-09-301-906-3 x Y91963

Align seg 1/1 to: Y91963 from: 1 to: 598

5359 TTTGAAGTGGTTAAACC.....GGGACCTCTGAGA 5390  
|||||.....  
117 PheGluValIleThrLeuAlaAsnThrLeuGlyPheSerPheTh 133  
5391 TGGCGTGGTGAGTACTTGAATCTGGCTATAGGCATTGAGAGGACAT 5440  
|||||.....GluSerT 144  
133 rGlyMetAsnSerPheLeuAspIle.....  
5441 ACAGGGGTTGTTAGTGGCTAGAAATATTCGGTCACTACCGCGGAAGGT 5490  
|||||.....  
144 yrArgThrLeuValArg.....ArgValSerGlySerAspGly 156  
5491 GTTCTGAACCTACCTATCAAGTTTATGAATCACTACCGGGCTTTCAC... 5538  
|||.....  
157 TyrSerAlaValLeuGluAlaLeuCysHisSerLeuHisGlyLeuArgAl 173  
5539 .....GTTTACAAGTCGGGCACAGATCTCATTT 5566  
|||.....  
173 aGluValAspValLeuAspLysValValLysGlnGluIleProLeuIleG 190  
5567 TTCAT.....TCAACACAGACGGCTTCGGGTG 5595  
|||.....  
190 lyHisLysHisGluLeuPheCysLysAsnValGlnGluLeuAspArgVal 206  
5596 AGACACCTACCG..... 5607  
::: |||  
207 LysTyrLysProAspGlyPheHisTyrTyrAsnValGluLeuSerAsnII 223  
5608 .TACGTACTCATA...GCTGAAAAGGTATCTTTTACCAGGGCAAGATG 5653  
||| |||.....  
223 eTyrGlyLeuValAsnThrHisLysLeuValTyrAsnAsn..... 236  
5654 TCAGACGGGTGGTAGTTTGGGGACAATCTGTCGTATCGGACGATATA 5703  
||| |||.....  
237 ..AspProValValLysGluSerGluGlyIleValLeuLeuGluPro... 251  
5704 CTGGTTTCCACGAT...GCCATTATTTGATAGTGCACTG..... 5742

252 .....HisGluIleSerPheAsnLeuIleArgSerLeuAlaLeuII 265  
|||||.....  
5743 .....AAAGTCGCTCGATGGTCGGCGAATCGTTTA 5779  
|||||.....ValGluValAsnL 278  
265 eAspLeuLeuValLysValSer..... 278  
5780 AGTCCTTCGAA...TATAAGTGTATATGCTCCCGAGTGGCGGTAAG 5826  
|||||.....  
278 ysAlaIleAspAsnValLysPheValAsnAlaValProGlyAlaGlyLys 294  
5827 ACGACGAGCTTAGTGGAGCAATTCGTT.....AAGTC 5858  
||| .....  
295 ThrTyrGlnIleLysGlnArgMetLeuArgTrpPheAspSerGluLysAs 311  
5859 ACCCAATAGCACACCCACCTATAGCGCTAATGTGGGAAGTTCTCAGGACA 5908  
::: |||.....  
311 pGlySerAlaLeuLeuValLeuThrSerSerArgAsnSerAlaAspThrL 328  
5909 TAAAT...ATGGCGGTGAAGAAGACAGATCCGAATTTGGGAAGGTCCTCAAC 5955  
::: |||.....  
328 euLysAlaPheAlaGlnGluLysArg.....LeuGlyLysMetIle 341  
5956 AGTCCTACACAGTTAACTCCAGGTCGTAAACTTTATCGTCAGGGGA... 6003  
::: |||.....  
342 GlnIleLeuThrValAspAlaPheLeu.....PheGlnAlaArgGlyAr 356  
6004 .....ATGTATATAAGGGTTTGGTGGATGAGGTGCACATGATGC 6043  
::: |||.....  
356 gAsnValArgLeuTyrLysThrLeuLeuIleAspGluCysTyrMetThrH 373  
6044 ATCAAGGCTTACTACACTAGGCGCTCTCCGACACCGCGCGTCGGAAGGC 6093  
|| |||.....  
373 isAlaGlyIleLeuArgGlyIleIleAlaValLysProGluGluCys 389  
6094 CTCCTTTTTCGAGACATAATCAGATACCATTCATAACACAGGGAAGGT 6143  
::: |||.....  
390 ValLeuTyrGlyAspArgArgGlnValProPheIleAsnArgIleLysLe 406  
6144 GTTAGGATGGATTGCTGTTTGTTCCTCAAGCAAGCAAGCGCTGTAT 6193  
::: |||.....  
406 uLeuAsnAspAsnLysSerPheLeuLysProSerLeuGlyAsnTyrSerG 423  
6194 ACACCTTCTAAATCGTACAGGTGCTCGTTAGATGTTTGTCTACTGTTGTC 6243  
::: |||.....  
423 luMetLeuIleThrArgArgCysProAlaAspIleCysTrpArgMetSer 439  
6244 TCAATGACCGTAAAGGGAACGGAAGTGTACCTGAAAGGTCGTTAG 6293  
::: |||.....  
440 Asn.....ValAsnAsnGlyLysLys.....GlyAspArgLeuTyrSe 452  
6294 CGGT.....AAGGACAAACCACTAGTAGATCGCTGT 6325  
||| .....  
452 rGlyProValLysLeuPheThrGlnSerLysProValLeuLysSerValT 469  
6326 CCAAAAGGCCAATGGAAACCACTGATGACGTAGCTGAATAAACGCTGAC 6375  
::: |||.....  
469 hrCysLysAlaPheSerLysGlyAspHisAsnLeuPheSerGlnValAsp 485  
6376 GTGTACTTGTCCATGACCCAGTTGGCAGAAAGTCGGATATGAAGAGTCGTT 6425  
::: |||.....  
486 ArgValMetThrPheThrGlnAsnGluLysAsnGluLeuIleSerGluTy 502  
6426 GAAGGAAAAAGGA.....AAAGAAACCACTAGTGA 6457  
::: |||.....  
502 rMetSerArgGlyIleGlyThrIleGlnAspAlaLysThrLeuIleGlyT 519  
6458 CAGTGCATGAAGCACACAGGAAAAACATTCAGTGTATGTTGTTTGG 6507  
||| |||.....  
519 hrValAlaGluSerGlnGlyGluThrTyrLysArgValHisLeuValAla 535  
6508 ACGAAGAAACCGGATGACTCCCTATTTCATTAACACACCGCATATCTGT 6557  
||| |||.....



```

343 luGlyTyrHisHisSerTyrGlnThrAlaValArgArgLysValLeu 359
||||:||||| |||
2236 GTGACAGGATCTATTCCTACGGCTCAACACTATAGGACATTCGAGAA 2285
||||:|||||
360 ValAspLysLysValLeuThrArgValThrGluValAlaPheArgGlnPh 376
2286 TAGAACGTTTGGATATGCCGTTGGCGGTCAGGTCCGAAGACC.... 2331
||| ||| |||:||||| |||
376 eArgProAsnAlaAspAlaHisSerAlaIleGlnSerIleAlaThrMetL 393
2232 .....CATGTCATTACAGGAGTCGCGTTGTCCACAGC 2364
|||:||||| |||: |||
393 euSerSerSerThrAsnHisThrIleIleGlyGlyValThrLeuIleSer 409
2365 AAGCTGATATTCCTCTGATGATATGTGGGTTTGTAGTCTGCTGTTAT 2414
:||||| |||
410 GlyLysProLeuSerProAspAspTyrIleProValAlaThrThrIleTy 426
2415 GGCTCAGCGCATTAAG..... 2430
||||: |||
426 rTyrArgValLysLysLeuTyrAsnAlaIleProGluMetLeuSerLeuL 443
2431 ..CATAGGGGAAGATATTCGCTCTATTAACCTTATAAAGCCAGTGAG 2478
||||:|||||
443 euAspLysGlyGluArgLeuSerThrAspAlaValLeuLysGlySerGlu 459
2479 GGGAGTCTCCCGGGGTCTTCAAGCTCTTCTTTCAGACCCTAGCGGATTG 2528
||| |||
460 GlyProThr.....TyrTyrSerGlyProThr 468
2529 TTTTTCGACGAGTCTCCGCTCTATCTAGGCAATGGTCACGATTAAC 2578
||| |||:||||| |||
468 rPheLeuSerAlaLeuAsp.....LysValAsnValProGlyAspP 482
2579 TCAACGTTTGGAGAGCTTATGCTATGCCAGAGCGTTTCATCCGTAA 2628
|| ||| |||:||||| |||: |||
482 he...ValAlaLysAlaLeuLeuSerLeuProLys..... 492
2629 GTACCTGGGTCTGTTGTTTACCATTTCGACTTCTCGAGCTTCAGACAG 2678
492 ..... 492
2679 GTTGGAGCTCAGGGTCCCTTTGATATTCGAAGGACACTTCGGTAGGA 2728
|||
493 .....ArgA 494
2729 AACTGAAGATAGTCGCTTCGCGCTCTCTAGGCTATCGTGGAGAT 2778
||||:|||||
494 sPheLysSer.....LeuPheSerArgSer..... 502
2779 TCAATTAAAGTCATGAAGCAATGAACACAGAAGATGGAAACCCCTGCC 2828
|||: |||
503 .....AlaThrSerHisSerGluArgThrPr 511
2829 AATTACTGAAGATCTGTATATCGTTTCATATGGGAAGCTTCTTAACG 2878
|||: |||
511 oValGlnAspGlu.....SerProV 518
2879 TCCACTGTACAGGCGCAGGTCTTCTTGGCGTTTCGAAGCGACCGTGGTT 2928
|||: |||
518 alArgCys..... 520
2929 TCGAGTCTTCTTAAGGTTTGTAGTCTGTTGGGCTGCCAGGAAGGCCTT 2978
520 ..... 520
2979 TCTGCGATTACGCTCTCTTTTCCACAGGTTCACTATTCTACAGCCGG 3028
||| |||:|||||
521 .....ThrAspGlyValPheTyr..... 526
3029 GTTAACTGNAGATGAAGGCTTGATGCTCTGGTGGCCACAGAGATGCT 3078

```

```

526 ..... 526
3079 ATAACCTCACCGGTGGCATACTACTGAGACGTCGCGCTAGCTGTGACAA 3128
||||: |||
527 .....ProIleArgMetLeu.....LeuLy 533
3129 GGTGCTAGCTGGACACGAAGAATTTTGGAGTGAAGTTTCTTAATGACT 3178
| :|||: |||
533 scysLeuGlySerAspLys.....PheGluSerValThrIleThrAsp. 547
3179 TCACCACCTTTCGTATTCCGGAATAAGTGCTTATCCGGGATATTCTCGTGGC 3228
547 ..... 547
3229 TCTTTGGGTGGGCCCAATTCATGATGAAGTATAGGCGCGAATTGCGGC 3278
547 ..... 547
3279 TAAGCTAGAAGGTACGCGGGCAGTAGTTACGAACCTCTAAGCTCGTTAA 3328
547 ..... 547
3329 GTTCACAAGCGCGGTGTTTACGCGTTTTAACTCTCTAGCACAGTATCC 3378
|||: |||: |||: |||: |||: |||
548 .....ProArgSerAsnThrGluThrThrValAsp 557
3379 GGTGGATCTTTAGTCTGCGAAGAGGCTTTTCGTCGGCGGTGACCGTCAC 3428
: |||: |||: |||: |||: |||
558 .....LeuTyrHisSerPheGlnLysLysIle..... 566
3429 TAGGGCGACCGTAGCTAAACGTCAAGTCCCTTAGCGTTGCTATCTGTTT 3478
566 ..... 566
3479 CTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACAT 3528
566 ..... 566
3529 GCTCTTCCACGGCACCTTAATGTTTCTTTGGTTTAGGGACATTTGCTTGG 3578
: |||: ||| ||| |||: |||
567 .....GluThrValPheSerPheIleLeuGlyLysIleAspG1 579
3579 GCGAGGGCTAGCGCGAATACTTGGAACTTGGAGGCTTCTCCAAATAATT 3628
|
579 yPro..... 580
3629 GGTGCGCTGTTCCGAGGTTTGTGGCGAGGAAGAGTGTCAAGCTCATTG 3678
580 ..... 580
3679 TTAGTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGTCTAA 3728
580 ..... 580
3729 CGACACCATACCTCAACTTCCTTACGTCGCCCGGTAGAGGGGAGGAATG 3778
|||: |||: |||: |||
581 .....SerProLeuIleSerAspProV 588
3779 TGTACGATGAGACGCTTAGG .....TATTACCGGGACTTTGACTATGAC 3822
|||: |||: |||: |||: |||: |||
588 alTyrPheGlnSerLeuGluAspValTyrAlaGluTrp..... 601
3823 GAAGGTGCTGCTCCATCTGCGACTCAGCATGAAGCGGTTCCCGGTGACGA 3872
601 ..... 601
3873 TAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATG 3922
: |||: |||: |||: |||: |||: |||
602 .HisGlnGlyAsnAlaIleAspAlaSerAsnTyr..... 612
3923 TGGCGGACGTTGGGATTAGCACCAACGGGGAAGTTACTGTTGAAGAAGAG 3972
612 ..... 612

```

```
3973 ACCATTACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGTTGC 4022
612 .....
4023 CCGTCTCAGCTGTGGCGGAAGACAAGGTGATCGTCGGGTTCTGGTA 4072
612 .....
4073 CCGCTGACGCTATGGCTTTTGTGAAAGTGTGAAAAAGGTGTCGACGAT 4122
613 .....AlaArgThrLeuLeuAspAsp 619
4123 GTCTTTCACCAACAGTCTAGTGGGAAACGGCTCGTGAGGTTGAGTGG 4172
620 ILeArgLysGlnLysGluSerLeuArgAlaLysAlaLysGluValG1 636
4173 CGGCAAGGGTTGCTCCAGAAAGCGTCGTCGTGAGGGCCGACACAAG 4222
636 uAspAlaGlnLysLeuAsnArgAlaIleLeu..... 646
4223 AAAGGGAAGAGCTGCAGATGTTACACAGCACAAACCGCGTCAACGAA 4272
646 ..... 646
4273 GCGCACAGGAGCCAGTACAGTCCAGTCTTGTGAGTTCCGCCACAGGCTGA 4322
646 ..... 646
4323 TATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAGTGAAC 4372
647 .....GlnValHisAlaTyrLeuGluAlaHisP 656
4373 AAGAAGTACCATTTGGCGACTGTTTCGGGGCCGACCCCAATCTCGATGAG 4422
656 roAsp..... 657
4423 AAACCGGCCCAAGTGTWACGACTGCTGTGTGAAGATAATTGACAAGGG 4472
658 .....GlyArgLysIleGluGlyLeuG1 665
4473 CAAGCGGTGCGTCATGTGGTGTGAGAAACAGAGTACAAGTCGAGCAGC 4522
665 yLeuSerSerGlnPheIleAlaLysIleProGluLeuAlaIleProAlaP 682
4523 CCAACAGAGGAGTTGACGATCAATGAAGCAAGCCGCTAAACAGCTT 4572
682 roLysProLeuProGluPheGluLysAsnAlaGluThrGlyGluIleLeu 698
4573 TCATCTTTAGAACGTGTTCTCGGTGTCAGCTCGATGTGTACAAGCA 4622
699 .....ArgIleAsnProHisSerAs 705
4623 ACCGACTATGCCACCAGGTTCCTCAACGCATTTACCTTTGTCGATAACT 4672
705 pAlaIleLeu.....GluAlaIleAspTyrLeuLysSerT 717
4673 TCAAGGGAGGAGTCGGCTTTTCTCAAGCTGGGTGAGGGGTATACC 4722
717 hrSerAlaAsnSerIleIleThrLeuAsnLysLeuGly..... 729
4723 TATAATGTTGTTAGCCATCTTTTCATCAGGGTGGCCTCGTCCTTAGAGGA 4772
729 ..... 729
4773 TATCTTAACGGCAATTAACTACCCAAAGCGTCTTCGACCACCTGTTTAGTGC 4822
730 .....AspHisCys..... 733
4823 AGAAGTACAAGATGGTGGAGCGGTACCATTCACAGCTGATCAGGAGGAG 4872
733 InTrpThrThrLysGlyLeuAspValValTrpAlaGlyAspAspLysArg 749
```

---

```
4873 .....TGCTATCCATCAGATAAACCCCTATCTTACGGTCAATCTCGTGGG 4916
750 ArgAlaPheIleProLysLysAsn.....ThrTrpValG1 761
4917 GAAGCGAAACTTCTCGACTAAGTGCAGGAAGGGTGAAGGTCACTGGTCA 4966
761 yProThrAlaArgSerTyrProLeuAlaLysTyrGluArgAlaMet.... 776
4967 TAAACGTAGCTTCGGGTGACTATTCTTTCTTATGCCTTCGCGGTTTTCAAAGG 5016
777 .....SerLysAspGlyTyr..... 781
5017 ACCGACTTTCATTCAGTAAACTCCATCGACGAAGGCGCATCAGTTTGAC 5066
781 ..... 781
5067 GTTCAGGGCAACTCGCGCGCTCTTTGGTGTAGCAGGATGTTGCAGTTAG 5116
781 ..... 781
5117 CCGCGCGCGTGTGCGATGAGAGTCACCAGGTGTTCCAAACCAGCAACCA 5166
781 ..... 781
5167 CAGAGCCAAAGTGTCTACCAGAAATACACCAAAATCGGGGGCAAGGC 5216
782 .....ValThrLeuArgTrpAspGlyGluVa 790
5217 TCTATCTGAGGAAGTGTAGGAAAGTCAAGGGAGGTGACATACTCA 5266
790 lLeuAspAlaAsnCys.....ValArgSerLeuSerGlnTyrGluI 804
5267 TA...TGGTGGCAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313
804 leValPheValAspGln.....SerCysValLysAlaSerAla 816
5314 GATAATCCAGTATGGCTCTTGAACCTGACTACACCCCAATGACATTTGA 5363
817 GluArgIleIleProSerLeuGluLys..... 825
5364 ACTGGTTAAACCGGGAGCCTCTGAAGATCCCGTGTGGAGTACTTGAAGT 5413
825 ..... 825
5414 ATCTGCTATAGGCATTTGAGAGACATACAGGCGTGTGCTTATGGCTAGA 5463
826 .....AlaLeuGlyLeuGlu.....Ala 831
5464 AATATTGCGCTCACTACCCCGAAGGTGTCTG.....AAAGTACC 5504
832 HisPheSerValThrIleValAspGlyValAlaGlyCysGlyLysThrTh 848
5505 TAAT.....CAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAGT 5548
848 rAsnIleLysGlnIleAlaArgSer..... 856
5549 CGGCGACAGATCTCATTTTTCATTCAACACAAAGACGGCTTGCGTGTGAGA 5598
856 ..... 856
5599 GACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAA 5648
857 .....SerGlyG1 859
5649 AGATGTCGACGGGTGGTAGCTTTGGCGACAACTCTGTTGTTATGCGAGC 5698
859 nAspValAspLeuIle.....LeuThrSerAsnArgSerSerAlaAspG 874
5699 ATATACTGGTTTTTCCACGATGCCATTAAATTTGATAGTGCACCTGAAAGTC 5748
874 lLeuLysGluThrIleAspCysSerProLeuThrLysLeuHisTyrIle 890
5749 GCTCGATCGGCATGTTGGGCGAATCGTTTAAAGTCTCTCGAATATAAGTG 5798
```



18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgLeuTyrAspThrA 35  
1307 AATTTGGCATATGTTTCTCAAGGAGCGCAATGATTCGGCAAAACCC 1356  
35 laVal...AspGluPheAsnAlaArgAsp.....ArgArgPro 46  
1357 TCACATAGTGGCATGTGTTCTCGAAGCGCGGAAAGGGAGAGCTGAG 1406  
47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63  
1407 GGAACCTTTCCAGAGCTTTTCGATACAGTTCTCCGACTCGGTCAGAGTA 1456  
63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80  
1457 GTCACCCATTCGCTAATGCCATGCGAGCTGT...TTCAATGGAATCTTT 1503  
80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96  
1504 TCCAGGAGGTGTGTAATGTGTGCTTCTTCGATATTGGGGAGGCTTCAC 1553  
97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113  
1554 GTATCATGTCAAGCTGCCATGTGAATGTCATGTATGCAATCCAGTCC 1603  
113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130  
1604 TAGACCTTTAAAGATCTGAAGCGGACAATCAATGAG.....ATC 1641  
130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146  
1642 CTCCTTTCTTCC...ACAGCTGGGGAGGATTGCTAGCTGTCC..... 1680  
147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheL 163  
1681 .....AGTGACCTTCTAACTGAAGCGGCTTCAA 1708  
163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180  
1709 AGTCTGTGTTACTGTAGTCAGAACTGCAGAACTGC...GATTCTAGA 1755  
180 IsAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196  
1756 GCCGATGCGGGTTTTATGGTGGATGTACGATATATCCCGCAGCAGGT 1805  
197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213  
1806 ACCAGGCTATGGATAGAGAGGGTGGCTGTTTCGACATAGCTCTTA 1855  
213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230  
1856 TCTTCCCGTGGAGTGTGTGTACGGTAAAGTTTACTTGAAGAA 1905  
230 IsPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246  
1906 CTCGATAGCTTTGGTGAAGAGGAGGTGATTACCTGGCTACAATGTGG 1955  
247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263  
1956 TCAGTGTGGTGAGTG...TATGACATTCCTTCTACGTAACGCGGT 2002  
263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280  
2003 TTTTTCACCTTTTCTTATACGCACTTCGTCGGGAACGTGTTTAAGCTA 2052  
280 yrValCysLysThrTyrPheProAlaSerAsnArgGluValTyrMetLys 296  
2053 GAGTATGAGGATACCGTGTGTGTACCATCATCTCACTATGTGTAGGC 2102  
297 GluPhe..... 298  
2103 TCAGAACTGACCTGGAACTGAGGTACGTATAGGTGCTGGTCCGCTGT 2152

298 ..... 298  
2153 TCGTGGGCAATCGCTGTGTTCATACCTGTGTAGCTGGTCTAGTGTG 2202  
298 ..... 298  
2203 TCCTTTAAGACAATAGTCTCGATTTCGGACTTTGTGCACAGGATCTATTTC 2252  
299 .....LeuValThrArgValAsnTh 305  
2253 CTAC.....GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296  
305 rTrpPheCysLysPheSerArgIleAspThrPhe..... 316  
2297 ACTATGCCGTTGGGGCGGTCCAGTCCAAAAGACCCATGTCATTACACGG 2346  
317 .....LeuLeuTyr 319  
2347 AGTCGCGTTTCCACAGCAAGTTGTATATTCTCTGATGATATGTGGG 2396  
320 LysGlyValAlaHisLysGlyValAsp..... 328  
2397 TTTAGTTGCTGCTGTATGCGCTCAGCGGATTAAAGATAGCGCAAGAGTA 2446  
328 ..... 328  
2447 TTCGCTCTATAACTTTATAAAGCCAGTGAGGGAGTCTCCGCCGGGTC 2496  
329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341  
2497 TTCAAGCTCTTCTTTCAGACCGTAGCGGATTTGTTTTCGAACGCACTCTC 2546  
342 .....TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355  
2547 CGTCTATGCTAAGCAATGGTGCACGATAACTTCAACGTTTGGAGACCC 2596  
355 uLeuGluAspSerSerValAsnTyrTrpPheProLysMetArgAspM 372  
2597 TTATGCTATGCCAGAGGTTTCATCCGTAAGTACCTGGGCTGTGTTGT 2646  
372 etValIleValPro..... 376  
2647 GTTACCATTTCGACTTCTGGAGCTTCAGACAGTTTGGAGCTCAGGGGTC 2696  
377 .....Le 377  
2697 CTTTTCATATTTCAAGGAGACCTTCGGTAGGAACCTGAAGAATAGTCGT 2746  
377 uPheAspIleSerLeuGluThr.....SerLysArgThrArgL 390  
2747 TCGCGCTCTCTAGGGCTATCGTGAAGATTCAATT...AAGGTCATG 2793  
390 ysGluValLeu.....ValSerLysAspPheValTyrThrValLeu 403  
2794 AAGGCAATGAAGACAGAGATGGAACCCCTCCCAATTTACTGAAGATTC 2843  
404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu.....ThrTyrSerAs 418  
2844 TGTATATCGGTTTCATAATGGGGAACGTTTCTAACGTCCTCACTGTACAGGG 2893  
418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430  
2894 CAGGTCTTCTTCGGGCTTCGAAAGCGACCGTGTTCGAGTGCTTTCAAG 2943  
431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445  
2944 GGTTTGGTAGCTCGTGGGCTCGCAGCAAGGCGCTTTTCTGGCATTACGTC 2993  
446 SerLeuLeuGlnSerLeuSerMetThr..... 454  
2994 GTTCTTTTCCACAGGTTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043  
455 .PhePhe..... 456

[illegible][illegible]



```
739 rglnLysPheGly..... 743
4876 TATCCATCAGATAACCCATCTTTGACGGTCAATCTCGTGGGAAGGCATA 4925
743 ..... 743
4926 CTTCGACTAAGTCAGGAAGGGTGAAGGTTCATGCTATGCTGATAAAGCTAG 4975
744 .....ValLeuAspValA 748
4976 CTTCCGGGTCACTATATTTCTTATGCTTGC.....GGT 5007
748 IaSerLysArgTrpLeuValLysProSerAlaLysAsnHisAlaTrpGly 764
5008 TTTCAAGGAGGCACCTTCATTCAGTAAC.....TCCATCGA 5045
765 ValValGluThrHisAlaArgLysTyrHisValAlaLeuLeuGluHisAs 781
5046 CGAAGGGCGCATCAGTTTGACGTTTCAGGCAACTCGGCGCGTCTTTGGTG 5095
781 pGlu.....PheGlyI 785
5096 TAGGCAGGATGTTGCAGTTAGCCGGCGCGTCTCGGATGACAAGTCACCA 5145
785 le..... 785
5146 GGTGTTCCAAACCACACACCACAGAGCCCAAGGTGTACCAGACAATCAC 5195
785 ..... 785
5196 ACCAAATCGGGGGCAAGGCTCTATCTCAGGGAAGTGTAGGGAAGTCA 5245
785 ..... 785
5246 AGGGAGGTGCGACATACGATATGCTGATGTCGCAACAAGATTACGTTAGGAAG 5295
786 .....IleThrCysAsp..... 789
5296 TGTGAGTGGCTCAGGGCTGATAATCCAGTCAGTGCTTGTACCTGACTA 5345
790 ...AsnTrpArg..... 793
5346 CACCCCAATGACATTTGAAGTGTAAACCGGACCTCTGAAGATGCGC 5395
794 .....ValAlaValSerSerGluSerV 801
5396 TCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTTGAGAGCACATACAGG 5445
801 alVal.....TyrSerAspMetAlaLysLeuArgThrLeuArg 813
5446 GCGTTGCTTATGGCTAGAAATATTCGCCCTCACTACCGCGCGAGGTGTTCT 5495
814 ArgLeuLeuLysAspGlyGluProHisValSerSerAlaLysValValLe 830
5496 GAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTACGTTTACA 5545
830 u..... 830
5546 ACTCGGGCACAGATCTCATTTTTTTCATTCACACACAGACGGCTTGGCTGTG 5595
830 ..... 830
5596 AGAGACCTACCGTACGTACTCATAGCTGAAAAGGTATCTTTACCAAGGG 5645
830 ..... 830
5646 CAAAGATGTCACGGCGGTGGTAGCTTTGGGGACAATCTGTTCTGATGCG 5695
831 .....ValAspGlyVal..... 834
5696 ACGATATACTGGTTTCCACGATGCCATTAATTGATAGGTGCACCTGAAA 5745
```

```
834 ..... 834
5746 GTCGCTCGATCGGCATGGTGGCGGAATCGTTAAGTCCTTGAATATAA 5795
834 ..... 834
5796 GTGCTATAATGCTCCGCCAGGTGGGGTAAAGACGACGACGCTTAGTGGAG 5845
835 .....ProGlyCysGlyLys..... 839
5846 AATTCGTTAAGTCACCAATAGCACAGCACCACTTACGGCTAATGTGGGA 5895
840 .....ThrLysGluIleLeuSerArgValAsn 848
5896 AGTTCTGAGGACATAAATATGCGG.....GTGAA 5924
849 PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetI 865
5925 GAAGAGAGATCCGAATTTGGAAAGGTCTCAACAGTCTACACAGTTAACT 5974
865 eaArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882
5975 CCAGGTGGTA...AATTTATCGTC.....AGGGA 6003
882 alArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898
6004 ATGTATAAAGGGTTTTGGTGGATGAGTGCACATGATGCATCAAGGCTT 6053
899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyC 915
6054 ACTACAACATAGGCGTCTTCGCAACGCGCGTCCGGAAGGCCCTCTTTTGG 6103
915 sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932
6104 GAGACATAAATCAGATACCATTCATAACAGGGAGAGAGGTGTTTAGGATG 6153
932 lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948
6154 GATTGCTGCTGTTTTTGTCCAAAGAGGAAGCGTTGTATACACTTCTAA 6203
949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965
6204 ATCGTACAGGTGTCGGTTAGATGTTGTCTACTTGTGTCTCTCA..... 6246
965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982
6247 .....ATGACCGTAAGGGGAACGGAAGGTGTACCCCTGAAAAG 6285
982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998
6286 GTCGTTAGCGGT.....AAGGACAACCAAGTAGTAAGATCGCTGTCCAA 6329
999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysG 1015
6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACCGCTGACGCT 6379
1015 yLys.....I 1017
6380 ACTTGTGATGACCCAGTGTGGAGAGTCGGATATGAAGAGCTGTTGAAG 6429
1017 leLeuThrPheThrGln.....SerAsp...LysGluAlaLeuLeu 1029
6430 GGAAGGAAAAGAACACCACTGATGACAGTCATGAAGACAGCAGGAAA 6479
1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyG 1045
6480 AACATTCACTGATGTTGTTAGGACGAAGAACGCCGATGACTCC 6529
1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062
6530 TATTCACATAA...CAACGCGATATACTGTTGGTTGTTCGAGACACACA 6576
1062 leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078
```

6577 CGCTCACTGTTTATCGCGCT .....CTGAGCTCAAGTTT 6611  
 ::::||||| ||| :::: ||| ||| ::  
 1079 LysSerLeuLysTyrTyrThrValMetAspProLeuValSerIleI 1095  
 6612 GGAGCAT.....AAGTCGGCACATATATAGCGAC 6642  
 : ||| ::::||||| :::: ||| ::  
 1095 eAtgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT:W87724

seq\_documentation\_block:  
 ID W87724 standard; Protein; 2161 AA.

XX AC W87724;  
 XX XX  
 XX DT 26-APR-1999 (first entry)  
 XX DE Rupestris stem pitting associated virus RSPAV-1 replicase.  
 XX XX  
 XX KW RSPAV-1: grape; transgenic plant; disease resistance; replicase.  
 XX OS Rupestris stem pitting associated virus.  
 XX XX  
 XX FT Key Location/Qualifiers  
 FT Misc-difference 1931  
 FT /note= "encoded by AG, apparent 1 nucleotide  
 deletion in open reading frame"

XX XX  
 XX PN W09852964-A1.  
 XX PD 26-NOV-1998.  
 XX XX  
 XX PF 20-MAY-1998; 98MO-US10391.  
 XX PR 17-DEC-1997; 97US-0069902.  
 XX PR 20-MAY-1997; 97US-0047147.  
 XX XX  
 XX PA (CORR ) CORNELL RES FOUND INC.  
 XX XX  
 XX PI Gonsalves D, Meng B;  
 XX XX  
 XX DR WPI; 1999-045297/04.  
 XX DR N-PSDB; V99284.

XX PT Isolated proteins from Rupestris stem pitting-associated virus and  
 PT related nucleic acid - vectors, host cells and transgenic Vitis  
 PT cultivars that are resistant to the virus  
 XX XX  
 XX PS Claim 4; Page 18-25; 163pp; English.

XX CC This is the amino acid sequence of the replicase of Rupestris stem  
 CC pitting associated virus RSPAV-1. It was deduced from the  
 CC nucleotide sequence of an open reading frame (ORF1) identified  
 CC in the RSPAV-1 genome (see V99284). Proteins and polypeptides  
 CC of RSPAV (see W87724-38), and nucleic acids encoding them (see  
 CC V99284-86) are claimed, as are expression systems, transformed  
 CC host cells, transgenic grape (Vitis) scion or rootstock cultivars  
 CC comprising DNA encoding a replicase, coat protein, or a protein  
 CC of a triple gene block, and a method of imparting RSPAV resistance  
 CC to Vitis scion or rootstock cultivars by transformation with such  
 CC DNA.  
 XX XX

SQ Sequence 2161 AA;

alignment\_scores:  
 Quality: 290.00 Length: 142  
 Ratio: 2.736 Gaps: 3  
 Percent Similarity: 74.648 Percent Identity: 41.549

alignment\_block:

US-09-301-906-3 x W87724

Align seq 1/1 to: W87724 from: 1 to: 2161  
 4660 TTTGTCGATAACTTGAAGGAGGAGTCCGGTCTTTTCTCAAAGCTGGG 4709  
 ||| ::::||||| :::: ||| ||| ::  
 733 PhelysAspArgLeuArgGlyArgGlyAlaSerPheSerLysProG 749  
 : ||| ::::||||| :::: ||| ::  
 4710 T...GAGGGGTATACCTATATATGCTGGTAGCCATGTTTCATCAGGGTGGC 4756  
 I :::: ::::||||| :::: ||| ::  
 749 yIleSerCysHisSerTyrAsnGlyGlySerHisThrSerLeuGlyTrp 766  
 4757 CTGTCGCCCTAGAGATATCTTAACGGCAATTAACTACCCAGCGCTTTC 4806  
 ||| ::::||||| :::: ||| ::  
 766 roLysPheMetAspGlnIleLeuSerSerThrGlyGlyArgAsnTyrTyr 782  
 4807 GACCACCTGTTTGTAGTGCAGAAAGTACAAGATGGTGGAGCGGTACCATCCA 4856  
 :::: ||| :::: ||| :::: ||| ::  
 783 AsnSerCysLeuAlaGlnIleTyrGluGluAsnSerLysLeuAlaLeuH 799  
 4857 CGCTGATGACGAGAGTGTCTATCCATCAGATAACCCCTATCTTGACGGTCA 4906  
 I ||| :::: ||| :::: ||| ::  
 799 sLysAspAspGluSerCysTyrGluIleGlyHisLysValLeuThrValA 816  
 4907 ATCTCGTGGGGAAGCAAACTTCTCG...ACTAAGTCGAGCAAG..... 4947  
 ||| ::::||||| :::: ||| ::  
 816 snLeuIleGlySerAlaThrPheThrIleSerLysSerArgAsnLeuVal 832  
 4948 GGTGGTAAAGTTCATGGTCATAAAGTAGCTTGGGTGACTATTTTCTTAT 4997  
 ||| :::: ||| :::: ||| ::  
 833 GlyGlyAsnHisCysSerLeuThrIleGlyProAsnGluPheGluMe 849  
 4998 GCCTTGGCGGTTTCAAAGGACGACCTTGCATTCAGTAACTCCATCGACG 5047  
 ||| :::: ||| :::: ||| ::  
 849 tProArgGlyMetGlnCysAsnTyrPheHisGlyValSerAsnCysThrP 866  
 5048 AAGGGCGCATCAGTTTCACGTTTCAGG 5073  
 ||| :::: ||| :::: ||| ::  
 866 roGlyArgValSerLeuThrPheArg 874

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT:R06531

seq\_documentation\_block:  
 ID R06531 standard; protein; 992 AA.

XX AC R06531;  
 XX XX  
 XX DT 19-DEC-1990 (first entry)  
 XX DE Protein encoded by RNA 1 gene of CMV.  
 XX XX  
 XX KW CMV; RNA 1.  
 XX OS Cucurbit mosaic virus.  
 XX XX  
 XX PN JP02167080-A.  
 XX PD 27-JUN-1990.  
 XX PF 19-DEC-1988; 88JP-0320015.  
 XX XX  
 XX PR 19-DEC-1988; 88JP-0320015.  
 XX XX  
 XX PA (NOGY-) NOGYO SEIBUTSU IDEN.  
 XX XX  
 XX DR WPI; 1990-241950/32.  
 XX DR N-PSDB; Q05751.

XX PT Genome RNA 1 gene for prepn. of cucumber mosaic virus - codes  
 PT protein contg. specified sequence of 994 amino acids used in prepn.  
 PT of resistant plants.  
 XX XX

Claim 1; Page 1; 10pp; Japanese.

CC The sequence was deduced from the DNA sequence which was prepd.  
CC from RNA isolated from CMV-O virus. The gene can be used to  
CC prepare plants which are resistant to cucumber mosaic virus.

XX  
SQ Sequence 992 AA;

## alignment\_scores:

Quality: 274.50 Length: 659  
Ratio: 0.850 Gaps: 25  
Percent Similarity: 49.014 Percent Identity: 22.155

## alignment\_block:

US-09-301-906-3 x R06531 ..

Align seg 1/1 to: R06531 from: 1 to: 992

```
4696 TTCTCAAGCTCGGTGACGGGTATACCTATATATGCTGTAGCCATGTTTC 4745
      ::::::::::::::: ||| ::::::::::::::: |||
4743 TyrGluGlnLeuGlySerMetTyrAspMetTyrAsnAlaSerIleSe 450
      ::::::::::::::: ||| ::::::::::::::: |||
4746 ATCAGGTGGCTCGTCCCTAGAGATATCTTAACGGCAATTAGTACC 4795
      |::: ||| ::::::::::::::: PheAlaAlaLeuThrArgA 459
4750 rlys...trp.....PheAlaAlaLeuThrArgA 459
      ||||| |||
4796 CAAGCGTCTTCGACCACTGTTTGTAGTCAGAGTACAAGATGGGTGAGGC 4845
      ||||| |||
4799 rArgValPheSerAlaVal..... 467
      ||||| |||
4846 GTACATTCCAGCTGATGACAGAGAGTCTATCCATCAGATTAACCTAT 4895
      ||||| |||
4868 .....HisAlaLeuPheProThrLeuArgProArgGluLulysG 481
      ||||| |||
4896 CTTGACGTCATCTCGTGGGAGGCAACTCTCGACTAGTCAGG. 4944
      ::::::::::::::: ||| ::::::::::::::: |||
481 uPheLysLeuSerThrPheValThrPheAsnGluGluCysSerP 498
      ||||| ::::::::::::::: |||
4945 ..AAGGGTGGTAAAGTCATGTCATAAAGTCTCGGTGACATTTT 4992
      ||||| ::::::::::::::: |||
498 heAspGlyGlyGluLulysVal...IleSerAlaAlaVal 513
      ||||| ::::::::::::::: |||
4993 CTTATGCCTTGGCGTTTCAAAGGAGCGACTTGCATTCAGTAACTCAT 5042
      ||||| ::::::::::::::: |||
514 .....AlaThrGlnAlaVal 518
      ||||| ::::::::::::::: |||
5043 CGACGAAGGCGCATCAGTTTGACCTTCAGGCGCAACTCGGCGCTTTG 5092
      ::::::::::::::: ||| ::::::::::::::: |||
518 lThrAspGlyLysValLeuAlaAlaGlnLysAlaGluLysLeu..... 532
      ::::::::::::::: ||| ::::::::::::::: |||
5093 GTGTAGGAGGATGTTGCGGTAGCGCGCGGTCTCGCATCAGAAAGTCA 5142
      ::::::::::::::: ||| ::::::::::::::: |||
533 ..AlaGluLysLeuAlaGlnProValAspGluValSerAsp.....Ser 546
      ::::::::::::::: ||| ::::::::::::::: |||
5143 CCAGTGTGTCACCAACCAACCAACCAACCAACCAACCAACCAACCAAT 5192
      ||| ||||| ::::::::::::::: |||
547 ProGluValProSerThrProAsp..... 555
      ||||| ::::::::::::::: |||
5193 CACACCAAAATCGGGGCGCAAGGCTCTATCTGAGGGAAGTGTAGGAAG 5242
      ||||| ::::::::::::::: |||
555 ..... 555
      ||||| ::::::::::::::: |||
5243 TCAAGGGAGGTGCATACCTCGATATGTTGCGCAACAAAGATTACGTTAGG 5292
      ::::::::::::::: ||| ::::::::::::::: |||
556 .....AspThrAlaAspValCysGlyLysGluGlnGluValSer 568
      ::::::::::::::: ||| ::::::::::::::: |||
5293 AAGTGTGAGTGCCTCAGGCTGATATCCAGTATGCTCTTGAACCTCA 5342
      ::::::::::::::: ||| ::::::::::::::: |||
569 GluLeuAspSerLeuSerAlaGlnThr..... 577
      ||||| ::::::::::::::: |||
5343 CTACACCCCAATGACATTGTAAGTGGTAAACCGGACCTCTGAAGATG 5392
      ::::::::::::::: ||| ::::::::::::::: |||
578 .ArgSerProIleThr...ArgValAlaGluArgAlaThr.....A 590
```

```
5393 CGTCTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGCATAC 5442
      ||:::::||||| |||
590 laMetLeuGluTyrAlaAlaTyr..... 597
      ::::::::::::::: |||
5443 AGGGCGTTCCTTATGGCTAGAAATATTGCCGTCACCTACCGCCGAAGGTT 5492
      ::::::::::::::: ||||| :::::::::::::::
598 .....GluLysGlnLeuHisAspThrThrValSerAsnLe 609
      ::::::::::::::: |||
5493 TCTGAAAGTACCTAATCAAGTT.....TATGAATCACTAC 5527
      ::::::::::::::: |||
609 ulysArgIleTrpAsnMetAlaGlyGlyAspLysArgAsnSerLeu 626
      ::::::::::::::: |||
5528 CGGGC.....TTTCACGTTTACAAGTCGGGCACAGATCTC 5562
      ||| ::::::::::::::: |||
626 luGlyAsnLeuLysPheValPheAspThrThrPheThrValAspPromet 642
      ::::::::::::::: |||
5563 ATTTTTCATTCAACACAGACGGCTTGGCTGTGAGAGACCTACCGTACGT 5612
      ::::::::::::::: |||
643 ValAsnIleHisPheSerThrGlyArgTrpMetArgProValProGlu 659
      ::::::::::::::: |||
5613 ACTCATAGCTGAAAAAGGTATCTTTACCAAGGC.....AAAGATG 5653
      ::::::::::::::: |||
659 ylleValTyrSerValGlyTyrAsnGluArgGlyLeuGlyProLysSerA 676
      ::::::::::::::: |||
5654 TCGACGCGGTGTAGCTTTGGCGACAAATCTGTCGTATCGACGACATATA 5703
      ::::::::::::::: |||
676 spGlyGluLeuPheIleValAsnSerGluCysValIleCysAsnSer... 691
      ::::::::::::::: |||
5704 CTGGTTTCCAGATGCCATTAATTTGTAGTGTGACCTGAAAGTCGCTCG 5753
      ::::::::::::::: |||
692 .....GluSerLeuSerAlaValThrArgSerLeuGlnAlaPr 704
      ::::::::::::::: |||
5754 ATGCGCATGTTGGGGAATCGTTTAAAGCTCTCGAATATATAAGTGTATA 5803
      |||
704 oThrGlyThrIleSerGln.....ValA 712
      ::::::::::::::: |||
5804 ATGCTCCCCAGGTGGCGGTAAAGACGACGCTTAGTGGACGAATTCGTT 5853
      ::::::::::::::: |||
712 spGlyValAlaGlyCysGlyLysThrThrAlaIleLysSerIlePhe... 727
      ::::::::::::::: |||
5854 AGTCACCCATAGCACACGCCACCATTCAGGCTAATGTGGGAAGTTCTGA 5903
      ::::::::::::::: |||
728 ..GluProSerThrAspMetIleValThrAlaAsnLysLysSerAlaGl 743
      ::::::::::::::: |||
5904 GGACATAAATATGGCGTGAAGAAGAGATCCGAATTTTGGGAAGGTCCTCA 5953
      |||
743 nAspValArgMetAlaLeuPheLysSerSerAspSerLysGluAlaCysA 760
      ::::::::::::::: |||
5954 ACAGTGCTACACAGTTAACTCCAGGCGGTGTAACCTTTATCGTCAGGGGA 6003
      ::::::::::::::: |||
760 laPheValArgThrAlaAspSerValLeuLeuAsnGluCys.....Pro 774
      ::::::::::::::: |||
6004 ATGTATAAAGGTTTGTGGTATGAGTGCACATGATGCATCAAGGCTT 6053
      ::::::::::::::: |||
775 ThrValSerArgValLeuValAspGluValValLeuLeuHisPheGlyGl 791
      ::::::::::::::: |||
6054 ACTACACTAGCGCTCTTCGCAACCGCGCTCGGAAGGCCCTCTTTTGTG 6103
      |||
791 nLeuCysAlaValMetSerLysLeuLysAlaValArgAlaIleCysPheG 808
      ::::::::::::::: |||
6104 GAGACATAAATCAGATACCATTTCATAACAGGAGGAGAGGTTTGTAGGATG 6153
      ||||| ::::::::::::::: |||
808 lYAspSerGluGlnIleAlaPheSerSerArgAspAlaSerPheAspMet 824
      ::::::::::::::: |||
6154 GATTGTGCTGTTTGTTCCTCAAGAGGAAAGCGTTGTATACACTCTCTAA 6203
      ::::::::::::::: |||
825 ArgPheSerLysIleIleProAspGluThrSerAspAlaAspThr..... 839
      ::::::::::::::: |||
6204 ATCGTACAGGTGTCGTTAGATGTTTGTCTACTTGTGTCTCCTCAATGACCG 6253
      ::::::::::::::: |||
840 .ThrPheArgSerProGlnAspValValProLeuValArgLeuMetAlaT 856
      ::::::::::::::: |||
6254 TA.....AGGGGAACGAAAGTGTACCTGAAAGGTCGTT 6291
```

```

856 hrLysAlaLeuProLysGlyThrHisSerLysTyrThrLysTrpValSer 872
6292 AGCGGTAGGCAACACAGTAGTAGTGCCTGTCACAAAGGCAATGG 6341
873 GlnSerLys.....ValLysArgSerValThrSerArgSerIleAl 886
6342 AACCCACTGATGACCTACCTGAAATAAACCGTGCAGCTGTACTGTGCATGA 6391
886 aSerValThrLeuValAlaSerLeuAspSerArgPheTyrIleThrMet 903
6392 CCCAGTTGGAGAGTCCGATATGAAGAGGTGCTTGAAGGGAAGGAAAA 6441
903 hrGlnAlaAspLysAla.....SerLeuIleSerArgAlaLys 915
6442 GAA.....ACACCAGTGATGACAGTGCA 6464
916 GluMetAsnLeuProLysThrPheTrpAsnGluArgIleLysThrValH1 932
6465 TGAAGCAGGAGGAAACATTCAGTGATGCTGGTATTCGTTAGGACGAAGA 6514
932 sGluSerGlnGlyLeuSerGluAspHisValThrLeuValArgLeuLys 949
6515 AAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTGTTG 6564
949 erThrLysCysAspLeuPheLysGlnPheSerTyrCysLeuValAlaLeu 965
6565 TCGAGACACACAGCTCCTACCTGGTTAT 6591
966 ThrArgHisLysValThrPheArgTyr 974
```

seq\_name: /SDSI/gcgdata/geneseq/geneseq/AA1994.DAT:R49657

seq\_documentation\_block:

ID R49657 standard; Protein: 1704 AA.

AC R49657;

DT 12-SEP-1994 (first entry)

DE Sequence of Heliothis armigera replicase encoded by RNA 1.

KW HaSV; RNA 1; small RNA virus; replicase.

OS Heliothis armigera stunt virus.

PN WO9404660-A.

PD 03-MAR-1994.

PF 13-AUG-1993; 93WO-AU00411.

PR 14-AUG-1992; 92AU-0004081.

PR 08-JUL-1993; 93US-0089372.

PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.

PA (PACI-) PACIFIC SEEDS PTY LTD.

XX Christian PD, Gordon KHJ, Hanzlik TN;

PI WPI: 1994-083180/10.

DR N-PSDB; Q58522.

XX Small RNA virus capable of infecting insect species, e.g.

PT Heliothis - and transgenic plants contg. viral nucleic acid, for

PT protection against insect pests

XX Disclosure: Figure 1; 183pp; English.

XX The inventors claim a virus comprising a genome hybridisable with

CC the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are

CC those given in Figs 1 and 2 of the specification. Isolated proteins

CC or polypeptide prepn. of the proteins or polypeptides derivable

CC from the virus are also claimed.

CC H. armigera larvae were raised and viral RNA was extracted. The virus

CC RNAs were reverse transcribed into cDNA. Clone E3 represents 99.7%

CC or RNA 1.(hr236 contains about 88% of RNA 2.)The full length clone

CC of RNA 1 was completed using PCR. RNA 1 encodes a protein of mol.

CC wt. 187,000 which is regarded as the replicase in view of its AA

CC sequence similarity in certain limited regions to replicases of

CC of other RNA viruses. The apparent mol. wt. of this protein upon

CC in vitro translation of virus RNA and SDS-PAGE is 195,000. The

CC sequence given in Figure 1 is in the same sense as the viral

CC (positive-sense) RNA. There are other small ORFs at the 3' end,

CC corresp. to the proteins P11a, P11b and P14.

XX

SQ Sequence 1704 AA;

alignment\_scores:

Quality: 237.50 Length: 991

Ratio: 0.571 Gaps: 50

Percent Similarity: 41.978 Percent Identity: 21.191

alignment\_block:

US-09-301-906-3 x R49657 ..

Align seg 1/1 to: R49657 from: 1 to: 1704

4273 GGCACAGGAGGAGGACAGTCCAGTCTTGTGAGTTGCCACAGGCTGA 4322

54 GlyThrLeuProThrGlnHisProIleLeuAlaGlyHisGlnArgVa 70

4323 TATTCCAAAGGTCAACCCAGTCCGAGTACATGCTCAGAAAGAGTGAAC 4372

70 lAlaGluGluValLeuHisAsnPheAlaArgGlyArgSerThrValLeuG 87

4373 AAGAAGTACATGCGACTGTTTTCGGCGCCACGCCAATCGTCGATGAG 4422

87 lulleGlyPro.....SerLeuHisSerAlaLeuLysLeuHisGlyAla 101

4423 AAACCCGCCCCAGTGTACGACTCGTGGTGTGAAGATAATTGACAAGG 4472

102 ProAsnAlaProAlaAlaAspTyrHisGlyCysThrLysTyrGlyThrAr 118

4473 CAAGGCGCTCGCTCATGTGCT.....GAGAAAAACAGGTACAGTCG 4516

118 gAspGlySerArgHisIleThrAlaLeuGluSerArgSerValAlaThrG 135

4517 AGCAGCCCAAA.....CAGAGGAGTTTGACGATCAATGAAGGCAAG 4557

135 lYargProGluPheLysAlaAspAlaSerLeuLeuAlaAsn...GlyIle 150

4558 GCCGTAACACAGCTTTGCATGTTTGAACCGTGTTCCTGC..... 4596

151 GlySerArgThrPheCysValAspGlyValGlySerCysAlaPheLysSe 167

4597 .....GGTGTGCGAGCTGGATGTGTACACGAGGAGCCTATCCGCCACCA 4639

167 rArgValGlyIleAlaAsnHisSerLeuTyrAspValThrLeu...GluG 183

4640 GGTTCCTCAACGCATTTACCTTTGTCGATACCTTGAAGGAGGAGGTCG 4689

183 luleuAlaAsnAlaPheGlu..... 189

4690 GTCTTTTCTCAAAGCTGGGTGAGGGGTATACCTATATATGTGTAGTACCA 4739

190 ..... 194

4740 TGTTTCATCAGGTCCTGCTGCCCTAGAGGATATCTTA..... 4779

194 sMetValArgAlaPheMetHisMetProGluGluLeuLeuTyrMetAspA 211

4780 .....ACGCCAATTAAGTACCCA 4797

211 snValValAsnAlaGluLeuGlyTyrArgPheHisValIleGluGluPro 227

4798 AGCGTCTCCACCACTGTTTACTGTCAGAAAGTACAGATGGGTGAGCCGT 4847  
::: ||| ||| ||||| ::  
228 MetAlaValLysAspCysAlaPheGln.....GlyGlyAspLe 240  
4848 ACCATTCCACGGCTGATCAGAGAGTGTCTATCCATCAGATACCCCTATCT 4897  
::: ||||| ::::: ||||| ::  
240 uArgLeuHisPheProGluLeuAspPheIleAsnGluSerGlnGluArgA 257  
4898 TGACGGCTCAATCTCGTGGGAGGCAAACTCTCGACTAAGTCAGCAGGAAG 4947  
::: ||||| ::::: ||||| ::  
257 rgIleGluArgLeuAlaIaArgGlySerTyrSerArgArgAla..... 271  
4948 GGTGTGTAAGCTCATGCTCAATAACCTAGCTCGGGTGAC..... 4986  
272 .....ValIlePheSerGlyAspAspTrpGlu 281  
4987 .....TATTTTCTTATG. 4998  
281 yAspAlaTyrLeuHisAspPheHisThrTrpLeuProTyrLeuValAla 298  
4999 .....CCTTCGGGTTT .....CAAAGG 5016  
298 rGAsnTyrProThrProPheGlyPheSerLeuHisIleGluValGlnArg 314  
5017 ACCCACTTCGATTTCAGTAAACTCCATCGACGAAAGGCGCATCAGTTTGAC 5066  
||| ||||| ::::: ||||| ::  
315 ArgHisGlySerSerIleGlu..... 321  
5067 GTTCAGGCAACTCGCGCGTCTTTGGTGTAGCGAGGATGTCGACTTAG 5116  
::: ||||| ||||| ||||| ||||| ::  
322 .LeuArgIleThrArgAlaProGly...AspArgMetLeuAlaValV 337  
5117 CC.....GCGCGGTGTCGGATG. 5134  
337 alProArgThrSerGlnGly. LeuCysArgIleProAsnIlePheTyrTy 353  
5135 .....AGAAGTCACCAAGTGTTCCAAACCCAGCAACCA 5168  
::: ||||| ::::: ||||| ::  
353 rAlaAspAlaSerGlyThrGluHisLysThrIleLeuThrSerGlnHisL 370  
5169 GAGCCCAAGTGTACCAGAACCAATCACCAAAATCGGGGGCAAGGCTC 5218  
::: ||||| ||||| ::  
370 ysValAsnMetLeu. LeuAsn..... 376  
5219 TATCTGAGGGAAGTGTAGGGAAGTCAAGGGAGTCTGACATACGTATA 5268  
::: ||||| ||||| ||||| ||||| ::  
377 .....PheMetGlnThrArgProGluLysGluLeuValAspM 389  
5269 TG.....GTCCGAACAAGATTACGTTTAGGAAGTGTGA 5300  
||| ||||| ::::: ||||| ::  
389 eLThrValLeuMetSerPheAlaArgAlaArgLeuArg...AlaIleVal 404  
5301 GTGGCTCAGGCTGATATTCAGTATGCTCTTGAACCTGACTACACCC 5350  
||||| ||||| ::::: ||||| ::  
405 ValAla.....SerGlu.ValThrGluSerSerTrpAsnI 416  
5351 CAATGACATTTGAAGTGTAAACCGGACCTCTCAAGATCCGCTCGTG 5400  
::: ||||| ::::: ||||| ::  
416 leSerProAlaAspLeuValArgThrValValSer..... 427  
5401 GAGTACTTGAAGTATCTGCTATAGGCATATGAGAGACATACAGGCGCTT 5450  
||||| ||||| ||||| ||||| ::  
428 .....LeuTyrValLeuHisIleIleGluArgArgAlaAla... 440  
5451 GCTTATGGCTAGAAATATTTCGCGTCACTACCGCGGAAGGTGTTCTGAAAG 5500  
::: ||||| ||||| ||||| ::  
441 .....ValAlaValLysThrAlaLysAspValPheG 452  
5501 TACCTAATCAAGTTTATGAATCACTACCGGCTTTTACCGTTTACAGTGC 5550  
::: ||||| ||||| ||||| ::  
452 lylGluThrSerPheTrpGluSerLeuLys.....HisValPheGlySer 466

5551 .....GGCACAGATCTCATTTT...CATTC 5573  
||||| ||||| ::::: ||||| ::  
467 CysCysGlyLeuArgAsnLeuLysGlyThrAspValValPheThrLysAr 483  
5574 AACACAAGAGCGCTTCGGTGTGAGACACCTACCGTAGCTACTC..... 5616  
::: ||||| ||||| ::::: ||||| ::  
483 gValValAspLysTyrArgValHisSerLeuGlyAspIleIleCysAspV 500  
5616 ..... 5616  
500 alArgLeuSerProGluGlnValGlyPheLeuProSerArgValProPro 516  
5617 .....ATAGCTGAAAAAGG 5630  
517 AlaArgValPheHisAspArgGluGluLeuGluAlaLeuArgGluAlaGl 533  
5631 TATCTTTACC.....A 5641  
| ::::: ::  
533 yCysTyrAsnGluArgProValProSerThrProProValGluGluProG 550  
5642 AGGCAAAAGATGTCGAC..... 5658  
::||| ||||| ::||| ||||| ::  
550 IndlyPheAspAlaAspLeuTrpHisAlaThrAlaAlaSerLeuProGlu 566  
5659 .....GCGGTGTAGCTTTGGCGCAATCTGTCGTATGCGACGATAT 5702  
||| ::::: ||| ||||| ||||| ::  
567 TyrArgAlaThrLeuGlnAlaGlyLeuAsnThrAspValLysGlnLeuLy 583  
5703 ACTGGTTTCCACGATGCCATTAAATTGATAGTGCACTGAAAGTCGCTC 5752  
::: ||||| ::::: ||||| ::  
583 sIleThrLeuGluAsnAlaLeuLysThrIleAspGlyLeuThrLeuSerP 600  
5753 GATGC...GGCATGTTGGCGAATCGCTTTAAGTCCTTCGAATATAAGTGC 5799  
||| ::::: ||||| ::  
600 roValArgGlyLeu.....GluMet 606  
5800 TATAATGCTCCCCAGGTGGCGTAAGACGACGACGCTTAGTG..... 5841  
||| ::::: ||||| ::||| ||||| ||||| ::  
607 TyrGluGlyProGlySerGlyLysThrGlyThrLeuIleAlaLe 623  
5841 ..... 5841  
623 uGluAlaAlaGlyLysAlaLeuTyrValAlaProThrArgGluLeuA 640  
5842 .....GACGAATTCGTTAAGTACCCCAATAGCACAGCCACCAT 5880  
||| ::::: ||||| ||||| ::||| ::||| ::  
640 rgGluAlaMetAspArgArgIleLysProProSerAlaSerAlaThr... 655  
5881 ACGGCTAATCTGGGAAGTTCGTGAGGACATAAATATATGCGGTGAAGAAGAG 5930  
::: ||||| ::||| ::||| ::||| ::  
656 .....GlnHisValAlaLeuAlaIleLeuArgAr 665  
5931 AGATCGGAATTTGGAAGGTCTACAGTGTACCCAGCTTAACCTCCAGG 5980  
| ::||| ||||| ||||| ::||| ::||| ::  
665 g...AlaThrAlaGluGlyAlaProPheAlaThrValValIleAspGluC 681  
5981 TGTAAACTTTATCTCGTCAGGGGAATGTATAAAGGTTTTTGGTGGATGAG 6030  
||| ::||| ||||| ||||| ::  
681 ysPhe.....MetPheProLeuValTyrValAlaIle 691  
6031 GTGCACATGATGATCAAGGCTTACTACAACCTAGGCGTCTTCGCAACCGG 6080  
||||| ::||| ::||| ::  
692 ValHisAlaLeu.....SerPr 697  
6081 CGGTCGGAAGGCTCTTTTGGAGACATAAATCAGATACCATTCATAA 6130  
::: ||||| ::||| ::||| ::||| ::||| ::  
697 oSerSerArgIleValLeuValGlyAspValHisGlnIleGlyPheIleA 714  
6131 ACAGGGAGAGGTTGTTAGGATGGATTGCTGCTTTTGTTCCAANGAG 6180  
::: ||||| ::||| ::||| ::  
714 sp..... 714  
6181 GAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCGGTTA.....GA 6224

```
715 .....PheGlnGlyThrSerAlaAsnMetProLeuValArgAs 727
6225 TGTTCGTACTGTTGTCCTCAATGACCGTAAGGGGAAACGAAAGTGT 6274
727 pValValLysGlnCysArgArgThrPheAsnGlnThrLysArgCys. 743
6275 ACCGTGAAAGTCGTAGCGGTAGGACCAACACAGTAGTAAGATCGCTG 6324
744 .ProAlaAspValValAla.....ThrThrPhePheGlnSerLeu 756
6325 TCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAAC..... 6369
757 TyrProGlyCysThrThrThrSerGlyCysValAlaSerIleSerHisVa 773
6370 .GTGACGTGTAC.....TTGTGCATGACCCAGT 6397
773 lAlaProAspTyrArgAsnSerGlnAlaGlnThrLeuCysPheThrGlnG 790
6398 TGGAGAGTCGGATATCAACAGGTCGTTGAAGGGAAAGGAAAGAAACA 6447
790 luGluLysSer.....ArgHisGlyAlaGluGly..... 799
6448 CCAGTGATGACGTGATGATGACAGGAGGAAACATTCAGTGATGGT 6497
800 ...AlaMetThrValHisGluAlaGlnGlyArgThrPheAlaSerValII 815
6498 ATTG...TTTAGCAGCAAGAAACCGATGACTCCCTATTCACTAAACAAC 6544
815 eLeuHisTyrAsnGlySerThrAlaGluGlnLysLeuLeuAlaGluLysS 832
6545 CGCATATACCTGTTGTTGTCAGACACACAGCGCTACGCTTTATGCC 6594
832 erHisLeuLeuValGlyIleThrArgHisThrAsnHisLeu..... 845
6595 GCTCTGAGCTCAAAGTTGGACGATAAGTGGCGACATATATTAGCGAGC 6644
846 .....TyrIleArgAsp.. 849
6645 GTCACTCAATCAGTATCCGACGCTTGTCTTCACAGCTTCGCCCGCGCTG 6694
850 ....ProThrGlyAspIleGluArgGlnLeuAsnHisSerAlaLysAlaG 865
6695 GTTGTCTTCAGGTATA 6711
865 luValPheThrAspIle 870
```

seq\_name: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:Y31381

seq\_documentation\_block:

ID Y31381 standard; Protein; 1698 AA.

```
AC Y31381;
XX
XX
XX 12-OCT-1999 (first entry)
XX
XX HEV-US1 ORF1 protein.
XX
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW vaccine; passive immunisation.
XX
XX Hepatitis E virus.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 174
FT /label= unknown
FT /note= "encoded by ATR"
FT Misc-difference 363
FT /label= unknown
FT /note= "encoded by RTG"
FT Misc-difference 1088
FT /label= unknown
FT /note= "encoded by CRC"
```

```
FT Misc-difference 1131
FT /label= unknown
FT /note= "encoded by YTC"
FT Misc-difference 1217
FT /label= unknown
FT /note= "encoded by CRC"
FT Misc-difference 1389
FT /label= unknown
FT /note= "encoded by GWC"
XX
PN W0919732-Al.
XX
XX 22-APR-1999.
PD
XX
XX 15-OCT-1998; 98WO-US21941.
PF
XX
XX 15-OCT-1997; 97US-0061199.
PR
XX
XX (ABBO ) ABBOTT LAB.
PA
XX
XX Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
PI
XX
XX WPI: 1999-288017/24.
DR
XX
XX N-PSDB; Z00195.
PT
XX
XX Detection of United States isolates of hepatitis E virus
XX
XX Claim 6; Page 175-180; 260pp; English.
XX
XX The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunosassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US1 ORF1 protein.
XX
XX SQ Sequence 1698 AA;
```

```
alignment_scores:
Quality: 188.50 Length: 1860
Ratio: 0.274 Gaps: 79
Percent Similarity: 36.989 Percent Identity: 18.065

alignment_block:
US-09-301-906-3 x Y31381 ..
Align seg 1/1 to: Y31381 from: 1 to: 1698

1375 TTCTGAGCCGCGGAAAGGAGAGCTGAGGAACTTTTCAGAGCT 1424
|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||
28 PheLeuSerArgValGlnThrGluIleLeuIleAsnLeuMetGlnProAr 44
1425 TTCGATACACTCTCCGACTCGGTCCGAGGAGTAGTCACCCATTCGCTAATG 1474
|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||
44 gGlnLeuValPheArgProGluValLeuTrpAsnHisProIleGlnArgV 61
1475 CCATGCGGAGCTGTTCAATGGAATCTTTCCAGAGGTGTGTAATGTG 1524
|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||
61 alIleHisAsnGluLeuGluGlnTyrCysArgAlaArgAlaGlyArg... 76
1525 TGCTTCTTCGATATTGGGGAGGCTTCACGTATCATGTCAAAGTCGCCCA 1574
|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||
77 Cys...LeuGluValGlyAla.....HisProArgSerIleAs 88
1575 TGTGAAGTGTATGTCATGCAATCCAGTCCTAGACGTTAAGATGTGAAGC 1624
|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||
88 nAspAsnProAsnValLeuHisArgCysPhe...LeuArgProValGlyA 104
```

```
1625 GGAGAAATCAATGAGATCCCTTTCTTCCACAGCTGGGGAGATTGCTAC 1674
|| : : : : : |||
104 rgaspValGlnArgTptYrSerAlaProThrArgGlyProAlaAaasn 120
: : : : : ||| ||| : : :
1675 GTCCTCAGCTACCTTCTTAAGTAAAGCGGCTTCAAAGCTCTGCTCTTAC 1724
: : : ||| : : : : : : : : : : : : : : : : : : : : :
121 CysArgArgSerAlaLeuArgGlyLeuProProAlaAaspArgThrTyrcy 137
: : : ||| : : : : : : : : : : : : : : : : : : : : :
1725 TAGTCGAGAATCGCAGAACTCGGATCTTAGAGCGGATCGGGGTTTT...A 1771
| : : : : : : : : : : : : : : : : : : : : : : : : : :
137 sphaespGlyPheSerArgcysAlaPheAlaGluThrGlyValAlaLu 154
: : : : : : : : : : : ||| : : : ||| ||| ||| ||| |||
1772 TGCTGATCTGACGATATATCCCGCAGCAGGTAGCAGAGGCTATGGAT 1821
: : : : : : : : : : ||| : : : ||| ||| ||| ||| |||
154 euTySerLeuHisAspLeuTptProAlaAaspValAlaGluAlaMetAla 170
: : : : : : : : : : ||| : : : ||| ||| ||| ||| |||
1822 AGAAGGGTCCGCTGGTTTTCAGACATAGCTCTTATGTTCGCCGTGGAGTT 1871
: : : ||| : : : : : ||| : : : ||| ||| ||| ||| |||
171 ArgHisGly***ThrArgLeuTyrrAlaAlaLeuHisLeuProGluVa 187
: : : ||| : : : : : ||| : : : ||| ||| ||| ||| |||
1872 GTGTACGCTAACGGTGAAGTTTACTTGAAGAAGCTCGATACGTTGGTGA 1921
: ||| : : : : : : : : : : : : : : : : : : : : :
187 LeuLeuProProGlyThrTyrrHisThrThrSerTyrr.....LeuLeuI 202
: : : : : : : : : : : : : : : : : : : : : : : : : :
1922 AGAGGAAGGTGAT.....TACCTGGCCACAAATGTTGCTCAGTGTGCT 1965
: : : : : ||| ||| : : : : : ||| : : : ||| : : : : :
202 leHisaspGlyAspArgAlaValValThrTyrrGlu...GlyAspThrSer 217
: : : : : ||| : : : : : ||| : : : ||| : : : : :
1966 GAGATGCTATGACATTCCTCTCAAGCTAAGCGGGTCTTTCACCTTTTC 2015
: : : ||| : : : : : ||| : : : ||| : : : : :
218 AlaGlyTyrrAsnHisAspValSerIleLeuArg.....Al 229
: : : : : : : : : : : : : : : : : : : : : : : : : :
2016 TTATGTACCACTTCGTCC.....GGGAACGTGTTTAAAGCTAGAGTAG 2059
: : : : : ||| : : : : : ||| : : : ||| : : : : :
229 atrpIleArgThrThrLysIleValGlyAsp...HisProLeuValIleG 245
: : : ||| : : : ||| ||| : : : : : ||| : : : ||| : : :
2060 AGGGATACCGTGTGT...GGTACCACATCACTCATGTGTAGGGCTCAG 2106
|| : : : : : ||| ||| : : : : : ||| : : : ||| : : :
* 245 luArgValArgAlaIleGlyCysHisPheValLeuLeuThrAlaAla 261
: : : : : : : : : : : : : : : : : : : : : : : : : :
2107 AAGTCACCT.....GGAACTGAGGT 2126
: : : ||| : : : ||| ||| : : : ||| ||| : : :
262 ProGluProSerProMetProTyrrValProTyrrProArgSerThrGluVa 278
: : : : : : : : : : : : : : : : : : : : : : : : : :
2127 TAGGTATAGTCGTGGTCCGCTGCTCGGGCAATCGCTGGTGTCCA 2176
| : : : : : ||| : : : ||| ||| : : : ||| : : : : :
278 lTyrrValArgSerIlePheGlyPro.....GlyGlySerProSerLeuP 293
: : : ||| : : : ||| ||| : : : ||| : : : ||| : : :
2177 TACCTGTTGTAGCTGGTCTCTAGTGTCTCTTTAAGACAATAGTCTCGAT 2226
: ||| : : : : : : : : : : : : : : : : : : : : :
293 heProSerAlaCysSerThrLysSerThrPheHisAlaValProVal... 308
: : : : : : : : : : : : : : : : : : : : : : : : : :
2227 TCGGACTTTGTCAGACAGATCTATTCCTACGGCGCTCAACACTATAGGAC 2276
: : : ||| : : : ||| ||| : : : ||| : : : ||| : : :
309 ...HisIleTrpAspArgLeuMetLeuPheGly.....AlaTh 320
: : : ||| : : : ||| ||| : : : ||| : : : ||| : : :
2277 ATTCGAGATAGACGTTT.....GAGTATCCCGTTGGGGCGGTACGT 2320
| : : : : : ||| : : : ||| : : : ||| : : : ||| : : :
320 rIeuAspaspGlnAlaPheCysCysSerArgIleuMetThrTyrrLeuArg 337
: : : : : : : : : : : : : : : : : : : : : : : : : :
2321 CGCAAAAGACCCATCTCATTCACAGAGCTCGCGTGTCTCCACAGCAAGGTT 2370
: : : : : ||| : : : ||| ||| : : : ||| : : : ||| : : :
337 lYleSerTyrrLysValThrValGlyAlaLeuValAlaAasnGluGlyTrp 353
: : : : : : : : : : : : : : : : : : : : : : : : : :
2371 GATATTTCTCCTGATCATATGTGGGTTTATGTTCGCT..... 2409
: : : ||| : : : ||| ||| : : : ||| : : : ||| : : :
354 AsnAlaSerGluAspAlaLeuThrAla***IleThrAlaAlaTyrrLeuTh 370
: : : : : : : : : : : : : : : : : : : : : : : : : :
2410 .....GTTATGCTCAGCGGATTAAGTAGGGCA 2440
: : : : : : : : : : : : : : : : : : : : : : : : : :
370 rIleCysHisGlnArgTyrrLeuArgThrGlnAlaIle.....SerL 384
: : : : : : : : : : : : : : : : : : : : : : : : : :
```



592 MetGlyAlaCylserHisSerLeuThrTrpTyrGluLeuThrProAlaGlyLe 608  
 |||||  
 ::::||||| ||| ::::|  
 3327 AAGTTCAACAGCGCCGGGTGTTTACGGGGTTTAACTCTAGCACAGTAT 3376  
 |::: :::: ||||| :|||  
 608 uGlnValArgIleSerSerAsnGlyLeuAuspCysThrAlaThrPheProp 625  
 CCGTGCGATCTTAGTCGTGCCAAGAGG.....TTTTCGTCTGC 3414  
 |||||::: :: ||| ||| |||  
 625 roGlyAlaProSerAlaAlaProGlyGluValAlaAlaPheCysSer 641  
 CGCGTCACCGTCACTAGGCGCACCTAGCTAAACGTCAAGTCCCCTTAGC 3464  
 |||:::  
 642 AlaLeu.....TyrAr 645  
 3465 GTTGCTATCGTTTTCTACCTCATACGCCCATTTCCGGCTGCAGTAGTTAG 3514  
 ||::: ::::|||||  
 645 gTyAsnArgPheThrGlnArgHisSerLeuThrGly.....G 658  
 GCATTGGCGCACATGCT.....CTT 3534  
 658 lYeuTrpLeuHisProGluglyLeuLeuGlyIlePheProProPheSer 674  
 CCACGCCACTTAATGTTT.....TCTTTGGTTTAGGGACATT 3572  
 ||| |||::: ||| ||| ||| ||| ||| |||  
 675 ProGlyHisIleTrpGluSerAlaAsnProPheCysGlyGluGlyThrLe 691  
 GCTTGGGGCGAGGCTAGCGGAATPACTTGGAAAGTTGGAGGCTCTCCA 3622  
 | ::::|||||::: ::::|||||  
 691 uTyr.....ThrArgThrTrpSerThrSerGlyPheSerS 703  
 ATAATTGGTGCCTGTTCCCGAGGTTGTTCCGCGAGGAAGAGTCTCAGC 3623  
 :::: .. 703 er..... 703  
 TCATTGTTACTGCCTATTACGCTAGGGCTATCTTTGTATCATAGGGGCTT 3722  
 703 ..... 703  
 3723 GCTTAACGACACCATACTCACTTGCTTAGCTCCACCGGTAGAGGGA 3772  
 703 ..... 703  
 3773 GGAATGTGTACGATCAGACGCTTAGGTATTACCGGGACTTTGACTATGAC 3822  
 704 .....AspPheSerProPro 708  
 GAAGTGTCTGGTCCATCTGGGACTCAGCATGAACGGTTCCTCCGGTGACGA 3872  
 ::::|||||::: |||||  
 709 GluAlaAlaAlaProAla.....MetAlaAlaThrProGlyLeuPr 722  
 3873 TAACGNATGGATCCACTTCTAGTGCTCAAGCTATGATGTGTGTACAATG 3922  
 :::: ||||| |||||::: |||:::  
 722 oHis.....SerThrProProValSerAspIleTrpValLeu..... 734  
 TCGCGCAGCTGGGGATTAGCACCAACGGGAAGTTACTGGTGAAGAAGAG 3972  
 734 ..... 734  
 3973 ACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTCG 4022  
 ||| ||| |||||  
 735 .....ProProProSer.....GluGluPheG1 742  
 4023 CCGGCTGCAGCTGTGGCGGAAGACAAAGGTGATCCCTCGGGTCTCGGTA 4072  
 ::::||||| :::: |||||:::|  
 742 nValAspAlaAlaProValProProAlaProAspProAlaGly..... 756  
 4073 CCGCTCACGCTATGCCTTTTGTGTGAAGTGTGAAAAAGGTGTGCGAGAT 4122  
 756 ..... 756  
 4123 GTCTTTTCACCAACAGCTCTAGTGGGGAACGGCTCGTGAGGTGAGGTGGA 4172

756 .....756

4173 CGGCAAGGGTTGTCTCCAGAAAGCGTCTCGGTGAGGCCGCACACAAG 4222  
||||| ||||| |||

757 .....LeuProGlyProValValLeuThrProPro.....766  
::: :::

4223 AAAGGGGAAGACTGCAGATGGTAACACAGCACAAACCGCGGTCAAACGA 4272

766 .....766

4273 GGCACAGGGAGCCAGTACAGTCCAGTCTTGAGTTGCCACAGGCTGA 4322  
||| : :

767 .....ProProProProValHI 772

4323 TATTCCAAGGTACCACCGTCCGAGGTACATGCTCAGAAAGAAGTGAAC 4372  
|||::: :

772 sLysProSerIle.....776

4373 AGAAGTACCATTGGCGACTGTTTCGGCGGCCACGCCAATCTCGATGAG 4422

776 .....776

4423 AACCSCCCCCAAGTGTACGACTCGTGGTGTGAAGATAATTGACNAAGG 4472  
||| |||||

777 ...ProProProSer.....780

4473 CAAGGCGTGCCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGC 4522

780 .....780

4523 CCAACAGAGGAGTTTACGATCAATGAAGCAAGCCGCTAAACAGCTT 4572  
::: ||| |||

781 ...ArgAsnArgArgLeu.....786

4573 TGCATGTTTAGAACGTGTTCTCGGCTGTCAGCTGGATGCTACNACGA 4622

786 .....786

4623 AGCGACTATCGCCACCAGGTCTCAAACGCACTTTACCTTTGTGATAACT 4672

786 .....786

4673 TGAAAGGAGAGTGCCTCTTTTCTCAAAGCTGGGTGAGGGGTATACC 4722  
||||| ||||| TyrThr

787 .....788

4723 TATAATGGGTAGCCATGTTTCATCAGGGTGGCTCTGTCCTAGANGA 4772  
||| ||::: ||| : : |||

789 TyrProaspGlyAlaLysValtyrAlagLy.....798

4773 TATCTTAACGGCAATTAAGTACCCAAGGCTCTTCGAC...CACTGT....4815  
|||::: ||::: ||| SerLeupheIusSerAspCysAspTr 807

4816 ..TTAGTGCAAG.....TACAAGATGGGTGAGCGGTACCA 4851  
|||::: : : : |||||:::

807 rpLeuValAsnAlaserAsnProGlyHisArgProGlyGlyGlyLeuCys 823

4852 .....TTCCACGTGATCAGCAGGAGTGCTATCATCATAGATAACCC 4892  
::: : : ||| |||||:::

824 HisAlaPheTyrGlnArgPheProGluAlaPheTyrProThrGlu....838

4893 TATCTTGACGGTCAATCTCGTGGGGAAGGCCAACTCTCGACTAAGTGA 4942

838 .....838

4943 GGAAGGOTGTAAGGTGATGGTATAAACGTAGCTTCGGGTGACTATTTT 4992  
|||

839 .....Phe 839

4993 CTATTGCTTGGGTTTTTCAAGACGCACCTTGCATTCAGTAACTCCAT 5042  
::: |||

840 IleMetArgGluGlyLeu.....845

```
5043 CGACGAAGGCGCATCAGTTTGACGTTACGGCAACCTCGCGCTTTG 5092
845 ..... 845
5093 GTGTAGGAGGATGTTGCAGTTAGCGGGCGGTGTCGGATGAGAATCA 5142
845 ..... 845
5143 CCAGGTGTTCCAAACCAACAGCAACAGCAAGGTGCTACCAACAAT 5192
846 .....AlaAlaIyrThrLe 850
5193 CACACCAAAATCGGGGGCAAGGCTCTATCTAGGGAAGTGTAGGAAG 5242
850 uThrProArg..... 853
5243 TCNAGGGAGGTCGACNACTCGATATGTCGCAACAGATTACGTTAGG 5292
853 ..... 853
5293 AAGTGTGAGTGGCTCAGGCGTCGATATCCAGTGATG...GCTCTTTGAACC 5339
854 .....ProIleIleHisAlaValAlaPr 861
5340 TGACTACACCCCAATGACATTTTGAAGTGGTTAAACCCGGGACCTCTGAAG 5389
861 oAspTyr..... 863
5390 ATGCCGCTGTGGAGTACTTGAGTATCTGGCTATAGGCATAGGAGGACA 5439
864 ..ArgValGluGlnAsnProLysArgLeu.....GluAlaAla 875
5440 TACAGGCGCTTCTTATGGCTACAAATATTCGGTCACTACCCGCGAAG 5489
876 TyrArgGluThrCysSerArgArgGlyThrAlaAla..... 887
5490 TGTCTGAAAGTCACTAATCAAGTTTATGAATCACTACCGGCTTTCAGG 5539
888 .....TyrProLeuLeuGlySerGlyI 895
5540 TTTACAGTCGGGCACAGATCTCATTTTTCATTCACACAGACGGCTTG 5589
895 leTyrGlnValProValSerLeuSerPheAspAlaTrpGluArgAsnHis 911
5590 CGTGTGAGAGCTACCGTACCTACTCATAGCTGAAAGAGTATCTTTAC 5639
912 ArgProGlyAspGluLeuTyrLeuThrGluProAlaAlaAsnTrpPheG 928
5640 CAAGGCAAGATGTCGACGCGGTGTAGCTTTGGCGCAACAATCTGTCG 5689
928 uAlaAsnLysProAlaGlnProValLeuThrIleThrGlu..... 941
5690 TAGCGACATATACTGTTTCCAGATGCCATTAATTTGATAGTGCA 5739
942 ....AspThrAlaArgThrAlaAsnLeuAlaLeuGluIleAspAlaAla 956
5740 CTGAAGTCGCTCGA...TGCGGCATGGTGGCGGAATCGTTTAAGTCCT 5786
957 ThrGluValGlyArgAlaCysAlaGlyCysThrIleSerProGlyIleVa 973
5787 CGAATATAGTGCTATATGTCCTCCCGAGTGGCGGTAAAGACGACGCT 5836
973 lHisTyrGlnPheThrAlaGlyValProGlySerGlyLysSerArgSerI 990
5837 TAGTGACGAATTCGTTAAGTACCCCAATAGCACGCCACCATTAACGCT 5886
990 le.....GlnGlnGlyAspValAspVal 997
5887 AATGTCGGAAGTCTCAGACACATAATATGGCGGTGAAGAAGAGATCC 5936
998 ValValValProThrArgGluLeuArgAsnSerTrpArgArg..... 1012
```

```
5937 GAATTTGGAAGGCTCTCAACACAGTGCTACCACAGTTAACTCCAGGGTGGTAA 5986
1013 .....GlyPheAlaAlaPheThrProHisThrAlaAlaArgValT 1026
5987 ACTTTTATCTCAGGGAATGTATAAAGAGGTTTGTGGTGGATGAGGTGCAC 6036
1026 hr.....IleGlyArgArgValIleAspGluAlaPro 1037
6037 ATGATGCATCAAGGCTTACTACAACCTAGGCTCTTCGCAACCGCGCGCTC 6086
1038 SerLeuProProHisLeuLeuLeuHisMet.....GlnArgAlaSe 1052
6087 GGAAGSCCTCTTTTGGACATATAATCAGATACCATTCATAAACAGGG 6136
1052 rSerValHisLeuLeuGlyAspProAsnGlnIleProAlaIleAspPheG 1069
6137 AGAAGGTGTTTAGGATGGATTGTGCTGTTTGTTCCTCCAAAG...AAGAA 6183
1069 luhis.....AlaGlyLeuValProAlaIleArgPro 1079
6184 AGCGTTGTATACACTTCT.....AAATCGTACAGGTGTCGTTAGA 6224
1080 GluLeuAlaProThrSerTrpTrp***ValThrHisArgCysProAlaAs 1096
6225 TGTCTGCTACTTGTGCTCTCAATGACCGTAAGGGAGACGAAAGTGT 6274
1096 pValCysGluLeu.....IleArgGly.....AlaT 1105
6275 ACCCTGAAAGGTGCTTAGCGGTAGGACAAACCACTAGTAAAGATCGCTG 6324
1105 yrProLysIleGlnThrThrSerArg.....ValLeuArgSerLeu 1118
6325 .....TCCAAAAGGCCAATTGSAACCACTGATGAGTACGTAGTAAATAA 6368
1119 PheTrpAsnGluProAlaIleGlyGlnLys..... 1128
6369 CGCTGACGTGTACTTGTGTCATGCCAGTTGGAGAAGTCGGATATGAAGA 6418
1129 .....LeuVal**ThrGln..... 1133
6419 GGTGCTTTGAAGGAAAGAAAGAACCACTGATGATGACGTGCATGAA 6468
1134 .....AlaAlaLysAlaAlaAsnProGlyAlaIleThrValHisGlu 1147
6469 GCACAGGAAACATTCAGTGTGTTGTTTGTAGCAGCAAGAAAGC 6518
1148 AlaGlnGlyAlaThrPheThrGluThrThrIleIleAlaThrAlaAspAl 1164
6519 CGATGACTCCCTATTCACTTAAACACCGCATATCTTGTGTTGTCGA 6568
1164 aArgGlyLeuIleGlnSerSerArgAlaHisAlaIleValAlaLeuThrA 1181
6569 GACACACAGCTCAGTGGTTTATGCGGCTCTGAGTCAAGTTGGACGAT 6618
1181 rgHisThrGluLysCysValIleLeuAspAlaProGlyLeuLeuArgGlu 1197
6619 AAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCATCATCCGACC 6668
1198 ...ValGly.....IleSerAspVa 1203
6669 TTTGCTTACAGGTTTCGCCCGCTGCT 6696
1203 lIleValAsnAsnPhePheLeuAlaGly 1212
```

```
seq_name: /SIBSL/gcgdata/geneseq/geneseqp/AA1999.DAT.Y31384
seq_documentation_block:
ID Y31384 standard; Protein; 1708 AA.
XX Y31384;
AC Y31384;
XX 12-OCT-1999 (first entry)
XX
```



2116 ....GGAACGTAGGTTACGTATAGG..... 2136  
:::|||||  
282 oArgSerThrGluValTyrValArgSerIlePheGlyProGlyGlySerP 299  
2137 ..TCGTTGTCGCGTCGTCGTCGGCAATCGCTGGTTC.....ATA 2178  
||||| ||||| :::: :::: :::: ::::  
299 roSerLeuPheProSerAlaCysSerThrLysSerThrPheHisAlaVal 315  
2179 CTTGTT...GTAGCTGGTTCTAGTGTGCTCTTAAAGACAATAGTCTCGA 2225  
||||| :::: :::: :::: :::: :::: :::: :::: ::::  
316 ProValHisIleTrpAsp\*\*\*LeuMetLeuPheGlyAlaThrLeu\*\*\*As 332  
2226 TTCGGACTTT...GTCACAGAGTATTCTCCACGCGCTCAACACTATAG 2272  
||||| :::: :::: :::: :::: :::: :::: :::: ::::  
332 pGlnAlaPheCysCysSerArgLeuMetThrTyr..... 343  
2273 GGACATTCAGAAATAGACGTTTGAGTAGCGCGTGGGGCGGTACGGTCG 2322  
:::|||||  
344 .....LeuArgGly 346  
2323 CAAAGACCCATGTCATTACAGGAGTCGCGTTGTCCACAGCAAGTTCA 2372  
:::|||||  
347 IleSerTyrLysValThrValGlyAlaLeuValAlaAsnGluGlyTrpAs 363  
2373 TATTTCTCTGATGATATGTGGGGTTAGTTGTCGCT..... 2409  
: ||| ||| :::: :::: :::: :::: :::: :::: ::::  
363 nAlaSerGluAspAlaLeuThrAlaValIleThrAlaAlaTyrLeuThrI 380  
2410 .....GTTATGGTCAGGCGATTAAAGGATAGGGCGAAG 2442  
::: ||| ||| :::: :::: :::: :::: :::: :::: ::::  
380 leCysHisGlnArgTyrLeuArgThrGlnAlaIleSerLysGlyMetArg 396  
2443 AGTATTCCG.....TCCTA 2456  
:::|||||  
397 ArgLeuGluValGluHisAlaGlnLysPheIleThrArgLeuTyrSerTr 413  
2457 TAACCTTTATAAAGCAGTACGAGGGAGTCTCGCGGGGTCTTCAAGCTCT 2506  
||| ||| :::: :::: :::: :::: :::: :::: ::::  
413 pLeuPheGluLysSerGlyArgAspTyrIleProGlyArgGlnLeuInP 430  
2507 TCTTTCAGACCGTAGCGAGTCTGTTTTCGACGAGCTCCGCTCTATGCT 2556  
::: ||| ||| :::: :::: :::: :::: :::: :::: ::::  
430 heTyrAlaGlnCysArgArgTyrLeuSerAlaGlyPheHisLeu\*\*\*Pro 446  
2557 AAGGCAATGGTCACAGATAACTCAACGTTTGGAGACGCTTATGCTCTAT 2606  
:::|||||  
447 Arg\*\*\*LeuValPheAsp.....GluSerValProCysAr 458  
2607 GCCCAGAGCGTTTCATCCGTAAGTACCTGGGTCTGT..TGTTGTTACCAATT 2655  
||||| :::: :::: :::: :::: :::: :::: :::: ::::  
458 gCysArgThrPheLeuLysLysValAlaGlyLysPheCysCysPhe.... 473  
2656 TGCACCTCTGGAGCTTCAGACAGGTGGAGCTCAGGGGTGCGCTTTGATAT 2705  
:::|||||  
474 .....MetArgTyrLeuGly...GlnGluCysThrCysPhe 484  
2706 TTCGAAGGAGACCTTCGGTAGGAACACTAGAAATAGTCCCTTCGCGGCT 2755  
:::|||||  
485 LeuGluPro.....AlaGluG 490  
2756 TCTTAGGCTATCTGGGAAGATTCAATTAAGTTCATGAAGCAATGAAG 2805  
||| ||| :::: :::: :::: :::: :::: :::: ::::  
490 yLeuValGlyAspGlnGlyHisAspAsnCluAlaTyrGluGlySerGlu. 506  
2806 ACAGAAGATGGAACCCCTGCGCAATTACTGAAGATTCTGTATATGCGTT 2855  
:::|||||  
507 ValAspProAlaGluProAlaHisLeuAspValSerGlyThrTyrAla... 522  
2856 CATAATGGGAACGTTTCTTAACGTCCACTGTACGAGGCGAGGCTCTCTGTG 2905  
|||||  
523 .....ValHis..... 524

2906 GCGGTTTCGAAAGCAGCACCGTGGTTTCAGTGTCTTAAGGTTTGGTAGCT 2955  
524 ..... 524  
2956 CGTGGGGCTCCGACGAGGCCCTTTCTGGCAATTACGTCGTTCTTTCCAC 3005  
524 ..... 524  
3006 AGGTTTCATTCTACGACCGCGGTTTAACTGAAGATGAAGGCTTGATG 3055  
525 .....GlyHisGlnLeuGluA 530  
3056 CTCTGTGCGCACAGAAATGTATAAAGTACCGGTGGGCATA...CTG 3102  
||||| ||| :::: :::: ||| |||  
530 leLeuTyrArg.....AlaLeuAsnValProHisAspIleAlaAla 543  
3103 GAGACATCCGCGGTA...GCTGTGACCAAGTCGTAGCTGGGAACGAAGA 3149  
:::|||||  
544 ArgAlaSerArgLeuThrAlaThrValGluLeuValAlaSerProAspAr 560  
3150 ATTTTGGAGTGAAGTTTCTTAATGACATTCACCACTTTTCGTATTGCGGA 3199  
:::|||||  
560 gluGluCysArgThrValLeuGlyAsnLysThr.....PheArgT 574  
3200 ATAAGGTGCTTATCGGGATATTTCGTGCGCTCTTTGGTGGCGGCCCAATT 3249  
::: ||| ||| :::: :::: ||| |||  
574 hrThrValValAspGlyAlaHisLeuGluAlaAsnGlyProGluGluTyr 590  
3250 GCATGGAAGTATAGCGCGGAATTGCGGCTAACGCTAGAGAGGTACCGGG 3299  
::: :::: ||| :::: :::: ||| |||  
591 valLeuSerPheAsp.....AlaSerArgGlnSerMetGlyAlaG 604  
3300 CAGT.....AGTTACGAA...ACTCTAAGCTCGTTAAGTTCACAAG 3337  
||||| :::: ||| ||| :::: :::: ||| |||  
604 ySerHisSerLeuTyrTrpGluLeuThrProAlaGlyLeuGlnValLysI 621  
3338 CGCGCGGTGTTTACGCGGTTTAACTCTAGCACAGTATCCGGTGGATCT 3387  
:::|||||  
621 leSerSerAsnGlyLeuAspCysThrAlaThrPhePro\*\*\*GlyGlyAla 637  
3388 TTAGTCTCCGGAAGAGG.....TTTTCTCGCGGTGACCGT 3425  
::: ||| ||| :::: :::: ||| |||  
638 ProSerAlaAlaProGlyGluVal\*\*\*AlaPheCysSerAlaLeu.... 652  
3426 CACTAGGCGACCGTAGCTAAAGCTCAAGTCCCTTAGCGTTGCTATCGT 3475  
653 .....TyrArgTyrAsnArgP 658  
3476 TTCTTACCTCATACGCCATTTCGGCTGCAGTATGTTAGGCATTTGGGCA 3525  
||||| :::: :::: ||| ||| :::: |||  
658 heThrGlnArgHisSerLeuThrGly.....GlyLeuTrpLeu 670  
3526 CATGCT.....CTTCCACGGCACTT 3545  
||| ||| ||| ||| ||| ||| |||  
671 HisProGluGlyLeuLeuGlyIlePheProProPheSerProGlyHisI 687  
3546 AATGTTT.....TTCTTTGGTTTACGGACATTCCTGGGGCGA 3583  
: ||| ||| ||| ||| ||| ||| |||  
687 eTrpGluSerAlaAsnProPheCysGlyGluGlyThrLeuTyr..... 701  
3584 GGGCTAGCGCGAATACTTGAAGTTCGAGGCTTCTCCAATAATTGGTGC 3633  
:::|||||  
702 .....ThrArgThrTrpSerThrSerGlyPheSerSer..... 712  
3634 GCTGTTCCCGAGGTGTTTGGGAGGGAAGAGTGTCAAGCTCATTTGTTACT 3683  
712 ..... 712  
3684 GCCTATTACGTAGGGGTATCTTTGATCATAGGGGCTTTCGTTAAGACA 3733  
712 ..... 712  
3734 CCATACCTCAACTTGCTTACGTCCACCAGGTAGAGGGAGGAATGTGTAC 3783

712 ..... 712  
3784 GATGAGCGCTTAGGTATTACCGGACTTTGACTATGACGAAGGTGCTGG 3833  
713 .....AspPheSerProProGluAlaAlaI 721  
3834 TCCATCTGGGACTCAGCATCAACCGCTCCCGGTGACGATAACGATGGAT 3883  
721 aProAlaSerAlaAlaAlaProGlyLeuPro.....TyrP 733  
3884 CCACCTTCTAGTGTCTCAAGCTATGATGTTGTTCACAAATGTGCCGACGTG 3933  
733 roThrProProValSerAspIleTyrValLeu..... 743  
3934 GGGATTAGCACCAACCGGGAAGTTACTGTTGAAGAAGACACCATTCACC 3983  
744 .....Pr 744  
3984 TCGAACCGTCCAAATACACTATATCTCGAGGAAGAGTTGCCCGCTCTGCAG 4033  
744 oProProSerGluGluSerHisValAlaAlaSerValProSerValP 761  
4034 CTGTGCGGAAACACAAGGTGATCCGTCCGGTTCGTGACCGCTGACGCT 4083  
761 ro.....GluProAlaGly..... 765  
4084 ATGGCTTTTGTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCA 4133  
765 ..... 765  
4134 ACAGTCTAGTGGGAACGCTCGTCTGAGTTGAGTGGACGGCAAGGGT 4183  
765 ..... 765  
4184 TGCTCCAGAAACGCTCGTCCGGTGTGAGCGCGCACACAAGAAAGGGAAGA 4233  
765 ..... 765  
4234 GGTGCAGATGGTAAACACAGCACAAACCGCGGTCAACGAAGGCACAGGA 4283  
765 ..... 765  
4284 GCCAGTACAGTCCAGTCTTGTGAGTTCCGCACAGGCTGATATCCAAAGG 4333  
766 ....LeuThrSerProIleValLeuThrProProProProProProv 781  
4334 TCACCCAGTCCGAGGTACATGCTCAGAAAGAGTGAACAAAGAGTACCA 4383  
781 al..... 781  
4384 TTGGCGACTGTTTTCGGGCGCCAGCGCAATCGTGATGAGAAACCCGCCCC 4433  
782 .....ArgLysProAlaIath 786  
4434 AGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGCGAAGCGCGTCG 4483  
786 rSer..... 787  
4484 CTCATGTGGCTGAGAAAAAACAGGTACAAAGTCGAGCAGCCCAACACAGAG 4533  
788 .....ProProProArgThrArg 793  
4534 AGTTTGACGATCAATGAAGCAAGCGCGGTAACACAGCTTTGCGATGTTAG 4583  
794 ArgLeuLeu..... 796  
4584 AACGTGTTCTCGGGTGTCCAGCTGGATGTGTACACGAAGCGACTATCG 4633  
796 ..... 796  
4634 CCACCAAGGTTCTCAACGCAATTTACCTTTTCGATAACTTTGAAGGAGG 4683

796 ..... 796  
4684 AGTGGCGTCTTTTCTCAAAAGCTGGGTGAGGGGTATACCTATAATGTGG 4733  
797 .....TyrThrTyrProAspG1 802  
4734 TAGCCATGTTTCATCAGGTTGGCTCGTGCCTAGAG....GATATCTTAA 4780  
802 yAlaLysValTyrAlaGlySerLeu\*\*\*GluSerAspCysAspTrpLeuV 819  
4781 CGCAATTAAGTACCCCAAGCGTCTTCGACCACCTGTTTAGTCGAGAAGTAC 4830  
819 aAlaSnAlaSerAsnProGly.....His 826  
4831 AGATGGGTGGAGCGGTACCATTCACGCTGATGACGAGGAGTGTATCC 4880  
827 ArgProGlyGlyGlyLeu...CysHisAlaPheTyrGlnArg...PhePr 841  
4881 ATCAGATAACCCCTATCTTGACGGTCAATCTCTGCGGGAAGCAAACTTCT 4930  
841 o.....GluAlaPheTyrS 846  
4931 CGACTAAGTGCAGGAAGGTGTTAAGTTCATGTCATAAAGCTAGCTTCG 4980  
846 erThrGlu..... 848  
4981 GGTGACTATTTCTTATGCTTTCGCGTTTTCAAAGGACGCACTTGCAATC 5030  
849 .....PheIleMetArgGluGlyLeu..... 855  
5031 AGTAAACTCCATCGACGAAGGCGCATCACTTTGACGTTTCAAGGCAACTC 5080  
855 ..... 855  
5081 GGCCTGCTTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGCGCGCTGTCG 5130  
855 ..... 855  
5131 GATGAGAAGTCAACAGGTGTTCCAAACAGCAACACAGAGCAAGGTGC 5180  
856 .....Al 856  
5181 TACCAGAACAAATCACACCAAAATCGGGGGCAAGGCTCTATCTGAGGAA 5230  
856 aAlaTyrThrLeuThrProArg..... 863  
5231 GTGCTAGGAAGTCAAGGGGAGGTGCGACATACGATATGTTGCGGAACAA 5280  
863 ..... 863  
5281 GATTACGTTAGGAAGTGTAGTGGCTCAGGGCTCATATCCAGTGTG... 5328  
864 .....ProIleIleH1 867  
5329 .GCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAACCG 5377  
867 sAlaValAlaProAspTyr..... 873  
5378 GGACCTCTGAAGATGCGCTGCTGGAGTACTTGAAGTATCTGGCTATAGGC 5427  
874 .....ArgValGluGlnAsnProLysArgLeu..... 882  
5428 ATTGAGAGGACATACAGGCGGTGCTTATGGCTAGAAATATTGCCGTCAC 5477  
883 ...GluAlaAlaTyrArgGluThrCysSerArgGlyThrAlaAla... 897  
5478 TACCGCGGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTAC 5527  
898 .....TyrProLeuL 901  
5528 CGGGCTTTTACAGGTTTACAGTCCGGCAGCATCTCATTTTTCATTCAACA 5577  
901 euGlySerGlyIleTyrGlnValProValSerLeuSerPheAspAlaIatrp 917

```
5578 CAAGACGGCTTCGCGTGTGAGAGACCTACGCTAGCTACTCATAGCTGAAAA 5627
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
918 GluArgAsnHisArgProGlyAspGluLeuTyrLeuThrGluProAlaAl 934
5628 AGGTATCTTTTACCAAGGGCAAGATGTGCGACCGGTGGTAGCTTTGGGG 5677
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
934 aAlatrPheGluAlaAsnLysProAlaGlnProAlaLeuThrIleThrG 951
5678 ACAATCTGTCGTATCGCAGCATATACCTGCTTTCCACCATGCCATTAAT 5727
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
951 lu.....AspThrAlaArgThrAlaAsnLeuAlaLeuGlu 962
5728 TTGATAGGTGCACTGAAAGTCGCTCGA...TGCGGCATGGTGGCGGAATC 5774
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
963 IlcAspAlaIaThrGluValGlyArgAlaCysAlaGlyCysThrIleSe 979
5775 GTTTAAGTCCTTCGAATATAAGTGTCTAATGCTCCCGCAGGTGGCGGTA 5824
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
979 rProGlyIleValHisTyrGlnPheThrAlaGlyValProGlySerGlyL 996
5825 AGACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCCAATAGCACGCC 5874
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
996 ySerArgSerIle.....GlnGlnGly 1003
5875 ACCATTAGCGCTAATGTGGAGATTCTGAGGACATAAATATGGCGTGAA 5924
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1004 AspValAspValValValProThrArgGluLeuArgAsnSerTrpAr 1020
5925 GAAGAGATCGCAATTTGGAAGTCTCAACAGTGTCTACCAAGTAACT 5974
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1020 gArgArg.....GlyPheAlaAlaPheThrProHisThrA 1032
5975 CCAGCGTGTAAACTTTATCGTCAGGGAATGTATAAAGCGTTTGGTG 6024
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1032 IaAlaArgValThr.....IleGlyArgArgValValIle 1043
6025 GATGAGTGCACATGATGCATCAAGCTTACTACAACTAGGCGTCTCGC 6074
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1044 AspGluAlaProSerLeuProHisLeuLeuLeuLeuHisMet..... 1058
6075 AACCGCGCGTGGGAAGCCCTCTTTTGGAGACATAATACAGATACCAT 6124
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1059 .GlnArgAlaSerSerValHisLeuLeuGlyAspProAsnGlnIleProA 1075
6125 TCATAACACAGGAGAGGTGTTAGATGGATTGCTGCTTTTGTTCCTCA 6174
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1075 IaIleAspPheGluHis.....AlaGlyLeuValPro 1085
6175 AAG...AAGGAAAGCGTTGTATACACTTCT.....AAATCGTACAG 6212
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1086 AlaIleArgProGluLeuAlaProHisSerTrpHisValThrHisAr 1102
6213 GTGTCGGTTAGATGTTGCTACTTGTTCCTCAATGACCGTAAAGGGAA 6262
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1102 gCysProAlaAspValCysGluLeu.....IleArgGly. 1113
6263 CGGAAAGTGTTCACCTGAAAGGTCGTTAGCGGTAAGGACAAACCCAGTA 6312
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1114 .....AlaTyrProLysIleGlnThrThrSerArg.....Val 1124
6313 GTAAGATCGCTG.....TCCAAAAGGCCAATTGGAACCACTGATGACGT 6356
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1125 IeuArgSerLeuPheTrpAsnGluProAlaIleGlyGlnLys..... 1138
6357 AGCTGAAATAACGCTGACGTGTACTTGTGCATGACCCAGTGGAGAAGT 6406
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1139 .....LeuValPheThrGln..... 1143
6407 CGGATATCAAGAGTTCGTTGAAGGAAAAGGAAAAGAACACCCAGTGATG 6456
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1144 .....AlaAlaLysAlaAlaAsnProGlyAlaIle 1153
```

```
6457 ACAGTGCATGAAGCACAGGAAAAACATTCAGTGATGTGGTATTGTTTAG 6506
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1154 ThrValHisGluAlaGlnGlyAlaThrPheThrGluThrIleIleAl 1170
6507 GACGAAGAAAGCCGATGACTCCCTATTCTACTAAACAACCGCATATACTTG 6556
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1170 aThrAlaAspAlaArgGlyLeuIleGlnSerSerArgAlaHisAlaIleV 1187
6557 TTGGTTTGTGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCA 6606
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1187 aIaLeuThrArgHisThrGluLysCysValIleLeuAspAlaProGly 1203
6607 AAGTTGGACGATAAGGTGCGGCACATATATTAGCGACGCTCACCTCAATC 6656
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1204 LeuLeuArgGlu...ValGly..... 1209
6657 AGTATCCGACGCTTTGCTTCACACGTTTCGCCCGGCTGGT 6696
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1210 .IleSerAspValIleValAsnAsnPhePheLeuAlaGly 1222
```



```
34 aLeuValSerIleThrThrSerGlyIleSerAspLysLeuLeuArg 50
2692 GGTGCTTTGATATTTCGAAGAGACCTTCGGTAGGAACACTGAAGATAG 2741
51 GlyAlaPheAspValSerLysLysAsnPheSerArgArgLeuArgSerSe 67
2742 TCGCTTTGGCGGCTTCTCTAGGGCTATCGTGGAAAGATCAATTAAGGTCA 2791
67 rArgLeuArgValPheSerArgAlaIleValGluAspThrIleLysValM 84
2792 TGAAGCAATGAAGACAGACAGATGAAACCCCTGCCAATTACTGAAGAT 2841
84 eLysGlyMetLysSerGluAspGlyLysProLeuProIleAlaGluAsp 100
2842 TCTGTATATGCTTCATATATGGGAACGTTTCTTAACGTCCACTGTACGAG 2891
101 SerValTyrAlaPheMetThrGlyAsnMetSerAsnValHisCysThrAr 117
2892 GCGAGTCTTCTTGGCGGTCGAAGCGCACCGCTGGTTTCGAGTGTTCVA 2941
117 gAlaGlyLeuLeuGlyGlySerLysAlaCysAlaAlaSerLeuAlaValL 134
2942 AGGGTTTGGTACTCGTGGGCTGCGACGAGCCCTTTCTCGCATTAGC 2991
134 ySGlyAlaAlaSerArgAlaThrGlyThrLysLeuPheSerGlyLeuThr 150
2992 TCGTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACCTGAAGA 3041
151 SerPheLeuSerAlaGlyLeuPheTyrAspGluGlyLeuThrProGl 167
3042 TGAAGGCTTGATGCTGTGTCGGCACAGAGAAATGCTATAAATCAACCGG 3091
167 yGluArgLeuAspAlaLeuThrArgArgGluHisAlaValAsnSerProV 184
3092 TGGGCTACTGAGAGCTGCGCGGTAGTCTGAGCAAGGTCTGACTGGA 3141
184 aIGlyLeuLeuGluProGlyAlaSerAlaLysArgValValSerGly 200
3142 ACGAAGAATTTTGGAGTCAAGTTTCCTTAAATGACTTCACCACTTCGT 3191
201 ThrLysAlaPheLeuSerGluLeuSerLeuGluAspPheThrThrPheVa 217
3192 ATTGCGGAATAAGTGCCTTATCCGGATATTCGTGGCGTCTTTGGTGGCG 3241
217 lIleLysAsnArgValLeuIleGlyValPheThrLeuSerMetAlaLeuT 234
3242 CCCCATTGCATGGAAGTATAGCGCGGAATTCGGCTAACCTAGACGG 3291
234 hrProValValTrpLysTyrArgArgAsnIleAlaArgThrGlyValAsp 250
3292 TACGCGGGCAGTAGTTACGAACCTCTAAGCTCGTTAAGTTCAACAGCGC 3341
251 ValPheHisArgAlaArgSerGlyThrAlaAlaIleGlyLeuGlnCysLe 267
3342 CGGTGCTTTACCGGTTTAACTCTAGCACAGTATCCGGTGGATCTTTAG 3391
267 uSerGlyGlyArgSerLeuAlaGlyAspAlaAlaArgGlyAlaLeuThrV 284
3392 TCGTGGGAAGAGGGTTTCGTGGCGGGTACCGTCACTAGGCGGACCGTA 3441
284 aIThrArgGlyGlyLeuSerSerAlaValAlaValThrArgAsnThrVal 300
3442 GCTAAACGTCAACTCCCTTACGCTTGTCTATCGTTTCTACCTCATACGC 3491
301 AlaArgArgGlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAl 317
3492 CATTTCCGGCTCCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGC 3541
317 aValSerGlyCysThrLeuLeuGlyIleTrpAlaHisAlaLeuProArgH 334
3542 ACTTAATGTTTCTTGGTTTAGGACATTCGTTGGCGGAGGCGCTAGC 3591
334 iSLeuMetPhePheGlyLeuGlyThrLeuPheGlyValSerAlaSer
```



```

4405 .....ACGCCAATCGTCGATGAGAAACCGCCCAAGTGT...AC 4442
618 GluValSerValProValAlaValGluLysThrValLeuSerValGluLys 634
4443 GACTCGTGTGTGAAGATAATTGACAAGGGCAAGCGCGTCGCTCATGTGG 4492
634 sAlaArgGluLeuLysAlaValAspLysGlyLysAlaValAlaHisAlaL 651
4493 CTGAGAAAAACAGGTACAAAGTCGAGCAGGCCCAACAGAGGAGTTTGCAG 4542
651 ySGLuValLysAsnValProValLysThrLeuProArgGlyAlaLeuLys 667
4543 ATCAATGAAGCAAGCGCGGTAAACAGCTTTGCGATGTTTGAACGCTTC 4592
668 ILeserGluAspThrValArgLysGluLeuLysCysMetPheArgThrCysSe 684
4593 CTCGGTGTGACGCTGGATGTGTACAAACGACCAAGCCACTATCCGCCACAGGT 4642
684 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 701
4643 TCTCAACGCAATTTACCTTTGTGCGATAACTTGAAGGGAGGAGTGGCGTC 4692
701 heSerAsnAlaPheThrPheValAspSerLeuLysGlyArgSerAlaVal 717
4693 TTTTCTCAAAGCTGGGTGAGGGGTATACCTATATATCGTGGTAGCCATGT 4742
718 PhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisVa 734
4743 TTCATCAGGTGGCCCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGT 4792
734 lSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysT 751
4793 ACCCAAGCGTCTCGACCACTGTTTGTGCGAGAAGTACAAGATGGGTGGA 4842
751 yrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 767
4843 GCGGTACCATTCACCGCTGATGACGAGGAGTGTATCCATCAGATACCC 4892
768 GlyValProPheHisAlaAspGluGlyCysTyrProSerAspAsnPr 784
4893 TATCTTCAGCGTCAATCTCTCGTGGGAGGCAAACTTCTCCACTAAGTGCA 4942
784 oIleLeuThrValIasnLeuValGlyLysAlaAsnPheSerThrLysCysA 801
4943 GGAAGGTGTGAAGTCATGCTCAACGTACGTCGCGGTGACTATTTT 4992
801 rGlyGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 817
4993 CTTATCGCTGCGGTTTTCAAAGGACGCACTTCGATTCAGTAACATCCAT 5042
818 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValIasnSerIl 834
5043 CGACGAAGGCGGCATCAGTTTACGTTTCAGGCAACTTCGCGCGCTCTTTG 5092
834 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArgValPheG 851
5093 GTGTAGCGAGGATGTTGCAAGTACCGCGCGCGCTGTCGGATGAGAAGTCA 5142
851 lValGlyArgMetLeuGlnLeuAlaGlyValSerAspGluLysSer 867
5143 CCAGGTGTTCCAAACAGCAACACAGAGCCAGGTCCTACCAAGACAAT 5192
868 ProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIl 884
5193 CACACAAAATCGGGGGGAAGCTCTATCTGAGGGAAGTGTGAGGGAAG 5242
884 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 901
5243 TCAAGGGGAGGTGCACATACCTCGATATGTCGCAACAAGATTACGTTAGG 5292
901 alLysGlyArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArg 917
5293 AGTGTGAGTGGCTCAGGCGTGATATCCAGTCATGGCTCTTGAACCTGA 5342

```

```

918 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuLysProGl 934
5343 CTACACCCCAATGACATTTTGAAGTGTAAACCGGACCTCTCAAGATG 5392
934 yTyrThrProMetThrPheGluValValLysAlaGlyThrSerGluAspa 951
5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTTAGAGGACATAC 5442
951 laValValGluTyrLeuLysTyrLeuAlaIleGlyIleGlyArgThrTyr 967
5443 AGGCGTTCCTTATGGCTAGAAATATTGGCTCACTACCCGCCGGAAGTGT 5492
968 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyVa 984
5493 TCTCAAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTTCACGTTT 5542
984 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1001
5543 ACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAAGACGGCTTCGCT 5592
1001 yLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArg 1017
5593 GTGAGAGACCTACCGTAGCTACTCATAGCTGAAAGAGGTATCTTTACCAA 5642
1018 ValArgAspLeuProTyrValPheIleAlaGluLysGlyIlePheIleLy 1034
5643 GGCAAAAGATGTCGACCGCTGGTAGCTTTGGCGCAACAATCTGTTCTGAT 5692
1034 sGlyLysAspValAspAlaValAlaLeuGlyAspAsnLeuSerValC 1051
5693 GCGACGATATACGTGGTTCACGATGCCATTAATTTGATAGGTGCACTG 5742
1051 ysAspIleLeuValPheHisAspAlaIleAsnLeuMetGlyAlaLeu 1067
5743 AAAGTCGCTCGATGCGCATGTTGGCGAATCGTTTAAAGTCCCTCGAATA 5792
1068 LysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTy 1084
5793 TAAGTGTCTAATCTCTCCCGCTGGCGTAAAGACGACGACGCTTAGTGG 5842
1084 rLysCysTyrAsnAlaProProGlyGlyLysThrThrMetLeuValAla 1101
5843 ACGAATTCGTTAAGTCACCAATAGCACAGCCACCATTCAGGCTAATGTG 5892
1101 spGluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1117
5893 GGAAGTCTCGAGCACATAAATATGGCGGTCAAGAGAGATCCGAATTT 5942
1118 GlySerSerGluAspIleAsnMetAlaValLysLysArgAspProAsnLe 1134
5943 GGAAGTCTCAACAGTCTACACAGTTAACTCCAGGTCGTAAACCTTA 5992
1134 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPheI 1151
5993 TCCTCAGGGGAATGTATAAAGGTTTGTGGATGAGGTGCACATGATG 6042
1151 leValArgGlyMetTyrLysArgValLeuValAspGluValTyrMetMet 1167
6043 CATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGCGCGCTCGGAAG 6092
1168 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGl 1184
6093 CCFCTTTTGGAGACATAAATCAGATACCATTCATAAACAGGAGAGAAG 6142
1184 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 1201
6143 TGTTTAGGATGGATGTGCTGTTTTTGTTCCTCAAGAGGAAAGCGTTGTA 6192
1201 alPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 1217
6193 TACACTCTAAATCGTACAGGTGTCGTTAGATGTTTGTCTACTGTTGTC 6242

```

```
1218 TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 1234
6243 CTCATGACCGTAAGGGGAACGAAAGTGTACCTGAAAGAGGTGCTTA 6292
|||||
1234 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 1251
6293 GCGGTAAAGGACAAACAGTAGTAAGATCCCTGTCCAAAGAGCCAAATTGA 6342
|||||
1251 exGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 1267
6343 ACCATGATGAGTAGCTAGCTGAATTAACGCTGACGTGTACTTGTGATGAC 6392
|||||
1268 ThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 1284
6393 CCAGTTGGAGAAGTCGGATATGAAGAGGTGCTGAAGGAAAGGAAAG 6442
|||||
1284 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLysG 1301
6443 AACACCACTGATGACAGTCAGTGAAGCACAGGGAACAAATTCAGTGAT 6492
|||||
1301 LuThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 1317
6493 GTGGTATTGTTAGCAGCAAGAAACCGGATGACTCCCTATTCACTAAACA 6542
|||||
1318 ValValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGl 1334
6543 ACCGATATACTTGTGGTTTGTGAGACACACACGCTCAGCTGTTTATG 6592
|||||
1334 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 1351
6593 CGCCTCTGACCTCAAAGTTGGACGATAAGTCCGGCACATATATTACGGAC 6642
|||||
1351 laAlaLeuSerSerGluLeuAspAspLysValGlyThrTyrIleSerAsp 1367
6643 GCGTCACCTCAATCAGTATCCACGCTTGTTCACACGTTGCCCGCGC 6692
|||||
1368 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 1384
6693 TGGTTGCTTCGAGGTATA 6711
|||||
1384 agLysCysPheArgGlyIle 1390
```

seq\_name: /cgn2\_6/ptodata/2/1aa/6\_COMB.pep:US-09-081-320-3

```
seq_documentation_block:
; Sequence 3, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Consalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESS: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-081-320-3

alignment_scores:
Quality: 290.00 Length: 142
Ratio: 2.736 Gaps: 3
Percent Similarity: 74.648 Percent Identity: 41.549

alignment_block:
US-09-301-906-3 x US-09-081-320-3 ..
Align seg 1/1 to: US-09-081-320-3 from: 1 to: 2161

4660 TTTGTCGATAACTTCAAAGGGAGGAGTGGCTTTTCTCAAAGCTGG 4709
|||||
733 PheLysAspArgLeuArgGlyArgGlyAlaSerPheSerLysProGl 749
4710 T...GAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGC 4756
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
749 yIleSerCysHisSerTyrAsnGlyGlySerHisThrSerLeuGlyTrpP 766
4757 CTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCAAGCGCTTC 4806
|||||
766 roLysPheMetAspGlnIleLeuSerSerThrGlyArgAsnTyrTyr 782
4807 GACCACGTGTTTAGTCGACAGGTACAAGATGGGTGGAGCGGTACCATCOA 4856
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
783 AsnSerCysLeuAlaGlnIleTyrGluGluAsnSerLysLeuAlaLeuH 799
4857 CGCTGATGACGAGGAGTGTATCCATCAGATAACCCCTATCTTCACGGTCA 4906
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
799 sLysAspAspGluSerCysTyrGluIleGlyHisLysValLeuThrValA 816
4907 ATCTCGTGGGGAAGCAAACTTCTCG...ACTAAGTCACAGGAAG..... 4947
|||||
816 snLeuIleGlySerAlaThrPheThrIleSerLysSerArgAsnLeuVal 832
4948 GGTGTTAAGTTCATGGTCATAAACGTAGTTCGGGTGACTATTTCTTTAT 4997
|||||
833 GlyGlyAsnHisCysSerLeuThrIleGlyProAsnGluPhePheGluMe 849
4998 GCCTTCGCGTTCCTCAAAGGACGCACTTCGATTCAGTAAACTCCATCGACG 5047
|||||
849 tProArgGlyMetGlnCysAsnTyrPheHisGlyValSerAsnCysThrP 866
5048 AAGGCGCATCATGTTTCAGCTTCAGG 5073
|||||
866 roGlyArgValSerLeuThrPheArg 874

seq_name: /cgn2_6/ptodata/2/1aa/6_COMB.pep:US-08-840-316-1
seq_documentation_block:
; Sequence 1, Application US/08840316
; Patent No. 6054567
; GENERAL INFORMATION:
; APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
; APPLICANT: Tsarev, Sergei. A., and Robinson, Robin A.
```

;; TITLE OF INVENTION: Recombinant Proteins Of  
;; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their  
;; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MORGAN & FINNEGAN  
;; STREET: 345 PARK AVENUE  
;; CITY: NEW YORK  
;; STATE: NEW YORK  
;; COUNTRY: USA  
;; ZIP: 10154  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: FLOPPY DISK  
;; COMPUTER: IBM PC COMPATIBLE  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/840,316  
;; FILING DATE: 11-APR-1997  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Richard W. Bork  
;; REGISTRATION NUMBER: 36,459  
;; REFERENCE/DOCKET NUMBER: 2026-4255  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 758-6800  
;; TELEFAX: (212) 751-6849  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1693 AMINO ACID RESIDUES  
;; TYPE: AMINO ACID  
;; STRANDEDNESS: UNKNOWN  
;; TOPOLOGY: UNKNOWN  
US-08-840-316-1

alignment\_scores:  
Quality: 179.50 Length: 888  
Ratio: 0.481 Gaps: 45  
Percent Similarity: 42.005 Percent Identity: 21.284

alignment\_block:  
US-09-301-906-3 x US-08-840-316-1 ..

Align seg 1/1 to: US-08-840-316-1 from: 1 to: 1693

```
4561 GGTAACAGCTTGGCATGTTTAGAAGCTGT.....TCCTGCGG 4598
|||||
425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaG1 441
4599 TGTGACGCTGCAT.....GTGTACACGAAGCG..... 4626
|||||
441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458
4627 .....ACTATCGCCACCATGCTCTCAACGCATTTACCTTTGTCGAT 4668
|||||
458 isCysArgThrAlaIleArgLysAlaValSerLysPheCysCysPheMet 474
4669 AACTGAAAGGGAGGAGTGGCTTTTCTCAAGCTGGGTGAGGGGTA 4718
|||
475 LysTrpLeuGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491
4719 T.....ACCTATAATGTT..... 4731
|||
491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspVala 508
4732 .....GGTAGCCATGTTTCATCA 4749
|||
508 sProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524
```

```
4750 GGGTGGGCTCGTGGCCCTAGAGATATCTTAACGGCAATTAAGTACCCCAAG 4799
|||
525 Gly.....ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539
|||
4800 CGTCTTCGACCACCTGTTTAGTGCAGAACTACAAGATGGGTGAGCGCTAC 4849
:
539 a.....GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552
:
4850 CATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCCTATCTG 4899
:
553 .....LysValSerGlnValAspGlyArgIleasp 562
4900 ACGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTCAGAGAGGG 4949
:
563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...Valas 578
:
4950 TGGTAAGTCTATGCTATAAACGTAGCTTCGGGTCACTATTTCTTATGC 4999
|||
578 pGlyAlaValLeuGluThrAsn..... 585
5000 CTTGCGGTTTTCAAGGACGCACCTTGCAT.....TCAGTAAACTCC 5040
|||
586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600
5041 ATCAGCAAGGCGGCATCAGTTTCAGCTTTCAGGCAACTCGGCGC..... 5085
:
601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617
5086 .....GTCCTTTGGTGTAGGCAGCATGTTGCAGTTAGCCG 5119
|||
617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634
5120 GCGCGTGTGCGATGAGAGTACCAGGTGTTCCTCA..... 5154
|||
634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650
5155 .....AACCAGCAACCCAG..... 5169
|||
651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667
5170 .....AGCCAAGTGTACCAGACAACTACACCAAAATCCG 5206
|||
667 nPheTrpPheHisProGluGlyLeuLeuGlyProPheAlaProPheSerP 684
5207 GGGCAAGGCTCTATCTGAGGGAAGTGTAGGGAAGTCAAGGGAGGTGCG 5256
|||
684 roGlyHisValTrpGlu.....SerAlaAsnPropheCysGlyGluSer 698
5257 ACATAC.....TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
|||
699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715
5301 GTGG.....CTCAGGGCTG 5314
|||
715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732
5315 ATAATCCAGTGATGGCTCTT.....GAACCTGACTACACC 5349
|||
732 laThrProThrProAlaAlaProLeuProProProAlaProAspProSer 748
5350 CCAATGACATTTGAAGTGGTAAACCGGACCTCTCGAAGATGCCGTCGT 5399
|||
749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764
5400 GGAGTACTTCAAGTATCTGCTATAGCCATTGAGAGGACATACAGGCGGT 5449
:
764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgLeuL 781
5450 TG.....CTTATGGCTAGAAATATTGCGGTC 5475
|||
781 eupheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797
```

```
5476 ACTACCGCCGAGGTTCTTGAAGTACTTAAT..... 5508
      ::::: ::::: |||
798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProG1 814
      ::::: ::::: |||
5509 .....CAAGTTATGAATCACTACCG..... 5529
      ::::: |||
814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831
      ::::: |||
5530 .....GGCTTTTCAGCTTTTACAAAGTCGGGCACA 5556
      ::::: |||
831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847
      ::::: |||
5557 GATCTCATTTTCATCAACACAGACGCTTGCCTGTGAGACACTACC 5606
      ::::: |||
848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864
      ::::: |||
5607 G..... 5607
      |
864 oLysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThra 881
      |
5608 .....TACGTACTCATAGCTGAAGAGGTATCTTTACCAAG.....GCC 5646
      ||| ::::: |||
881 laAlaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896
      ||| ::::: |||
5647 AAAGATGTCACCGCGGTGGTA.....GCTTTGGCGCACAACTGTT 5687
      ::::: |||
897 ProSerPheAspAlaTrpGluArgAsnHisArgProGlyAspGluLeuTy 913
      ::::: |||
5688 CGTA.....TGCG 5695
      ::::: |||
913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysp 930
      ::::: |||
5696 ACGATATACTGTTTCCACGAT.....GCCATT 5724
      ::::: |||
930 roThrLeuThrIleThrGluAspValAlaAlaArgThrAlaAsnLeuAlaIle 946
      ::::: |||
5725 AATTGATAGGTCCACTGAAGTCGCTCGATGC..... 5757
      ::::: |||
947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963
      ::::: |||
5758 .....GCCATGGTGGCGAATCGTTTAACTCCTCGAATATAAGTGT 5800
      ::::: |||
963 lThrProGlyValVal.....GlnTyrGlnPheT 973
      ::::: |||
5801 ATAATGCTCCCGCAGGTGGGTGAAGACGACGCTTAGTCGGACGAATTC 5850
      ::::: |||
973 hrAlaGlyValProGlySerGlyLysSerArgSerIle..... 985
      ::::: |||
5851 GTTAAGTACCCCAATAGCACAGCCACCAATTACGGCTAATGTGGGAAGTTC 5900
      ::::: |||
986 .....ThrGlnAlaAspValAspValValValValProth 997
      ::::: |||
5901 TGAGGCATAAATATGGCGGTGAAGNAGAGATCCGAATTTGGAAGGTC 5950
      ::::: |||
997 rArgGluLeuArgAsnAlaTrpArgArg.....Glyp 1009
      ::::: |||
5951 TCAACAGCTCTACACAGTTAACTCAGGTGTAACTTTATCGTCAGG 6000
      ::::: |||
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr..... 1021
      ::::: |||
6001 GGAATGTATAAAGGTTTTGGTGGATGAGGTGCACATGATCATCAAGG 6050
      ::::: |||
1022 ...GlnGlyArgArgValValIleAspGluAlaProSerLeuProProHi 1037
      ::::: |||
6051 CTTACTACAACTAGCGCTTCGCCACCGCGCGTCGGAGGCCTCTTTT 6100
      ::::: |||
1037 sLeuLeuLeuLeuHisMet.....GlnArgAlaAlaThrValHisLeuL 1052
      ::::: |||
6101 TTGGAGACATAAATCAGATACCATTCATAAACAGGAGGAAGGTGTTAGG 6150
      ::::: |||
1052 euGlyAspProAsnGlnIleProAlaIleAspPheGluHis..... 1065
      ::::: |||
6151 ATGGATTGTGCTGTTTTTGTCCAAAG...AAGGAAAGCGTTGTATACAC 6197
```

```
1066 .....AlaGlyLeuValProAlaIleArgProAspLeuAlaProth 1079
      ||| ::::: |||
6198 TTCT.....AATTCGTACAGGTGTCGGTTAGATGTTTGTCTACTTGT 6238
      ::::: |||
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095
      ::::: |||
6239 TGTCTCAATGACCGTAAGGGAAACGGAAGTGTTCACCTGAAAAGGTC 6288
      ::::: |||
1096 .....IleArgGly.....AlaTyrPrometIleGln 1104
      ::::: |||
6289 GTTAGCGGTAAAGACAAACACAGTAGTAAAGTCGCTG.....TCCAAAAG 6332
      ::::: |||
1105 ThrThrSerArg.....ValLeuArgSerLeuPheTrpGlyGluPr 1118
      ::::: |||
6333 GCCAATTGGAACCACTGATGACGTACTGAAATAAACGCTGACGTGACT 6382
      ::::: |||
1118 oAlaValGlyGlnLys.....L 1124
      ::::: |||
6383 TGTGCTATGACCCAGTTGGAGAGTCGATATGAAGAGTTCGTTGAAGGCA 6432
      ||| ::::: |||
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer..... 1137
      ::::: |||
6433 AAAGAAAAAGAACACCACTGATGACAGTCATGAAGACACAGGAAAAAC 6482
      ::::: |||
1138 .....ValThrValHisGluAlaGlnGlyAlaTh 1147
      ::::: |||
6483 ATTCAGTGTGTTGTTATTGTTAGGACGAAGACCGGATGACTCCCTAT 5532
      ::::: |||
1147 rTyrThrGluThrThrIleAlaThrAlaAspAlaArgGlyLeuIleG 1164
      ::::: |||
6533 TCACATAACAACCGCATATCTGTTGGTTGTCGAGACACACACGCTCA 6582
      ::::: |||
1164 InsSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180
      ::::: |||
6583 CTGCTTTATGCGCTCTGAGCTCAAAAGTTGGACGATAAGTCGGCACATA 6632
      ||| ::::: |||
1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly.... 1194
      ::::: |||
6633 TATTACGCGAGCGCTCACCTCAATCAGTATCCGACGCTTTGCTTCACAGT 6682
      ::::: |||
1195 .....IleSerAspAlaIleValAsnAsnP 1203
      ::::: |||
6683 TCGCCCGCGCTGCT 6696
      ||| ::::: |||
1203 hePheLeuAlaGly 1207

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US93-08849A-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9308849A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08849A
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US07/947,263
; FILING DATE: 18-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: William S. Feller
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4032 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; PCT-US93-08849A-1

alignment_scores:
  Quality: 179.50      Length: 888
  Ratio: 0.481         Gaps: 45
  Percent Similarity: 42.005      Percent Identity: 21.284

alignment_block:
US-09-301-906-3 x PCT-US93-08849A-1

Align seg 1/1 to: PCT-US93-08849A-1 from: 1 to: 1693

4561 GGTAACAGCTTTGCATGTTTGTAGAACGTGT.....TCCTCGCG 4598
|||||
425 GlyargGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaG1 441
4599 TGTGCACCTCGAT.....GTGTACACGAAGCG.....4626
|
441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458
4627 .....ACTATCGCCACGAGTTCTCAACACGATTTACCTTTGTGCAT 4668
|||||
458 isCysArgThrAlaIleArgGlyAlaValSerLysPheCysCysPheMet 474
4669 AACTTGAAGGAGGAGGTGGCGTCTTTTCTCAAGCTGGGTGAGGGCTA 4718
|||
475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491
4719 T.....ACCTATAATGGT.....4731
|||
491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508
4732 .....GGTAGCCATGTTTCATCA 4749
|||||
508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524
4750 GGTGGCGCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAG 4799
|||
525 Gly.....ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539
4800 CGTCTTCGACCACTGTTTGTGCAAGAGTACAAGATGGGTGGAGCGCTAC 4849
:
539 a.....GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552
4850 CATTCACCGCTGATGACGAGGAGTGTATCATCATCAATACCTACTTCTG 4899
:
553 .....LysValSerGlnValAspGlyArgIleAsp 562
4900 ACGGTCAATCTCGTGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGG 4949
:
563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578
4950 TGTAAAGGTGATGGTCATAAAGTAGCTTCGGGTGACTATTTCTTATGC 4999
|||
578 pGlyAlaValLeuGluThrAsn.....585
5000 CTTCGCGGTTTCAAGACGCCACTTTCAT.....TCAGTAACCTCC 5040

```

```

||| .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600
586 .....ATCCACGAAGGCGCATCAGTTTGCAGTTCAGGCAACTCGCGC..... 5085
5041 ATCCACGAAGGCGCATCAGTTTGCAGTTCAGGCAACTCGCGC..... 5085
||| .....MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617
601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617
5086 .....CTCTTTGTGTAGGACGAGTGTTCAGTTAGCGG 5119
||| .....uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634
617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634
5120 GCGCGTGTGGATGAGAGTACCAGGTGTTCCTCA.....5154
||||| .....roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650
634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650
5155 .....AACCAGCAACACACAG.....5169
||||| .....AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667
651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667
5170 .....AGCCAAGTGTACACAAATCACACCAAAATCGG 5206
||||| .....PheTrpPheHisProGluGlyLeuGlyProPheAlaProPheSerP 684
667 PheTrpPheHisProGluGlyLeuGlyProPheAlaProPheSerP 684
5207 GGGCAAGGCTCTATCTGAGGGAAGTGTAGGGAAGTCAAGGGAGGTCG 5256
||||| .....roGlyHisValTrpGlu.....SerAlaAsnProPheCysGlyGluSer 698
684 roGlyHisValTrpGlu.....SerAlaAsnProPheCysGlyGluSer 698
5257 ACATAC.....TCGATATGTCGCAACAGATTACGTTAGGAAGTGTGA 5300
||||| .....ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715
699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715
5301 GTGG.....CTCAGGGCTG 5314
||||| .....715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732
715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732
5315 ATATCCAGTGATGGCTCTT.....GAACCTGACTACACC 5349
||||| .....732 laThrProThrProAlaAlaProLeuProProAlaProAspProSer 748
732 laThrProThrProAlaAlaProLeuProProAlaProAspProSer 748
5350 CCAATGACATTGTAAGTGGTAAACCGGACCTCTGAAGATGCCGTCGT 5399
||||| .....749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764
749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764
5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGCGT 5449
||||| .....764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781
764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781
5450 TG.....CTTATGGCTAGAAATATTGCCGTC 5475
||| .....781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797
781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797
5476 ACTACCGCGGAAGGTGTTCTGAAAGTACCTAAAT.....5508
||||| .....798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814
798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814
5509 .....CAAGTTTATGAATCCTACCG.....5529
||||| .....814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831
814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831
5530 .....GGCTTTCAGCTTTACAAAGTCGGGCACA 5556
||||| .....831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847
831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847
5557 GATCTCATTTTTCATTCACACAGACGCGTGTGCTGTGAGACCTTACC 5606
||||| .....848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864
848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864
5607 G.....5607
| .....864 OlysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881
864 OlysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881
5608 .....TACGTACTCATAGCTGAAAAAGGTATCTTTACCAG.....GGC 5646
||| .....TACGTACTCATAGCTGAAAAAGGTATCTTTACCAG.....GGC 5646

```

```
881 laLaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896
5647 AAAGATCTCCAGCGGTGTA.....GCTTTGGCGCACAAATGTGT 5687
    ::::| | | | | | | | | | | | | | | | | | | | | | | |
897 ProSerPheAspAlaTrpGluArgAsnHisArgProGlyAspGluLeuTy 913
5688 CGTA.....TGGC 5695
    ::::| | | | | | | | | | | | | | | | | | | | | |
913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysP 930
5696 ACCATATCTGGTTTCCAGAT.....GCCATT 5724
    ::::| | | | | | | | | | | | | | | | | | | | | |
930 roThrLeuThrIleThrGluAspValAlaAlaArgThrAlaAsnLeuAla 946
5725 AATTTCATAGGTCCACTGAAGTCGCTCGATGC..... 5757
    ::::| | | | | | | | | | | | | | | | | | | | | |
947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963
5758 .....GGCATGGTGGCGAATCGTTTAAGTCCTCGAATATAAGTGT 5800
    ::::| | | | | | | | | | | | | | | | | | | | | |
963 lThrProGlyValVal.....GlnTyrGlnPheT 973
5801 ATAATGCTCCCGCGGTGCGTAAGCAGCAGCGTTAGTCGACGAATTC 5850
    ::::| | | | | | | | | | | | | | | | | | | | | |
973 hrAlaGlyValProGlySerGlyLysSerArgSerIle..... 985
5851 GTTAAGTCACCAATAGCAGCGCCACATAGCGTAAATGTGGGAAGTTC 5900
    ::::| | | | | | | | | | | | | | | | | | | | | |
986 .....ThrGlnAlaAspValaspValvalvalProTh 997
5901 TGAGGACATAATATGGCGGTGAAGAGAGAGATCGGAATTTGGAAGTTC 5950
    ::::| | | | | | | | | | | | | | | | | | | | | |
997 rArgGluLeuArgAsnAlaTrpArgArg.....GlyP 1009
5951 TCAACAGTCTACCACAGTTACTCAGGCGTAAACTTTATCGTCAGG 6000
    ::::| | | | | | | | | | | | | | | | | | | | | |
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr..... 1021
6001 GGAATGTATAAAGGTTTTGTGGTAGGTCACATGATCATCAAGG 6050
    ::::| | | | | | | | | | | | | | | | | | | | | |
1022 ...GlnGlyArgValvalleileaspGluAlaProSerLeuProProHi 1037
6051 CTACTACAACTAGGCGTCTCCCAACCGCGCGTCGGAGGCGCTCTTT 6100
    ::::| | | | | | | | | | | | | | | | | | | | | |
1037 sLeuLeuLeuLeuHisMet.....GlnArgAlaAlaThrvalHisLeuL 1052
6101 TTGCGACATAAATCAGATACCATTCATAAACAGGAGAGAGTGTAGG 6150
    ::::| | | | | | | | | | | | | | | | | | | | | |
1052 euGlyAspProasnGlnIleProAlaIleAspPheGluHis..... 1065
6151 ATGGATTGTCTCTTTTCTTCCAAAG...AAGGAAAGCGTTGTATACAC 6197
    ::::| | | | | | | | | | | | | | | | | | | | | |
1066 .....AlaGlyLeuValProAlaIleArgProAspLeuAlaProTh 1079
6198 TTCT.....AATCTGACAGTGTCCGTAGATGTTGCTACTTGT 6238
    ::::| | | | | | | | | | | | | | | | | | | | | |
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095
6239 TGTCTCAATGACCGTAAGCGCAACGGAAGTGTACCTCAAAAGGTC 6288
    ::::| | | | | | | | | | | | | | | | | | | | | |
1096 .....IleArgGly.....AlaTyrProMetIleGln 1104
6289 GTTAGGGTAAGCAACACAGTAGTAGTCCGCTG.....TCCAAAG 6332
    ::::| | | | | | | | | | | | | | | | | | | | | |
1105 ThrThrSerArg.....ValLeuArgSerLeuPheTrpGlyGluPr 1118
6333 GCCAATTGGAACACTGATGACGTAGCTAGTAATAAACGCTGACGTACT 6382
    ::::| | | | | | | | | | | | | | | | | | | | | |
1118 oAlaValGlyGlnLys.....L 1124
6383 TGTGATGACCCAGTTGGGAAGTCGGATATGAAGAGTCTGTGAAGGA 6432
    ::::| | | | | | | | | | | | | | | | | | | | | |
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer..... 1137
```

```
6433 AAAGAAAAAACACCAGTCATGACAGTCATGAAGACACAGGAAAAAC 6482
    ::::| | | | | | | | | | | | | | | | | | | | | |
1138 .....ValThrValHisGluAlaGlnGlyAlaTh 1147
6483 ATTCACTGATGGTATTTAGGACGAGAAAGCCGATGACTCCCTAT 6532
    ::::| | | | | | | | | | | | | | | | | | | | | |
1147 rTyrThrGluThrThrIleAlaThrAlaAspAlaArgGlyLeuIleG 1164
6533 TCACATAAACACCGCATATACCTTGTGTTGTCGAGACACACACGCTCA 6582
    ::::| | | | | | | | | | | | | | | | | | | | | |
1164 lnsSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180
6583 CTGGTTTATCGCGCTCTGAGCTCAAAAGTTGGACGATAAGTCGGCACATA 6632
    ::::| | | | | | | | | | | | | | | | | | | | | |
1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly..... 1194
6633 TATTAGCGACGCTCACCTCAATCATATCCGACGCTTTGCTTTCACACGT 6682
    ::::| | | | | | | | | | | | | | | | | | | | | |
1195 .....IleSerAspAlaIleValAlaAsnAsnP 1203
6683 TCGCCCGCGCTGCT 6696
    ::::| | | | | | | | | | | | | | | | | | | | | |
1203 hePheLeuAlaGly 1207
```

```
seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pcp: PCT-US93-08849-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9308849
; GENERAL INFORMATION:
; APPLICANT: Tsarev, Sergei A., Emerson,
; APPLICANT: Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08849
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,263
; FILING DATE: 18-SEP-1992
; NAME:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bork, Richard, W.
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; PCT-US93-08849-1
```

alignment\_scores:

Quality: 179.50 Length: 888  
 Ratio: 0.481 Gaps: 45  
 Percent Similarity: 42.005 Percent Identity: 21.284  
 alignment\_block:  
 US-09-301-906-3 x PCT-US93-08849-1 ..  
 Align seg 1/1 to: PCT-US93-08849-1 from: 1 to: 1693  
 4561 GGTAAACAGCTTTGCATGTTTAGAACCTGT .....TCCTCGCG 4598  
 |||:||||| ||| |||:|||||  
 425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaG1 441  
 4599 TGTCCACCTCGAT .....GTCTACAACGAACG ..... 4626  
 |||:||||| |||:||||| |||:|||||  
 441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458  
 4627 .....ACTATCGCCACACAGGTTCTCAACAGCATTTACCTTTGTCGAT 4668  
 |||:||||| |||:||||| |||:|||||  
 458 iCysArgTrpAlaIleArgLysAlaValSerLysPheCysCysPheMet 474  
 4669 AACTTGAAGGGAGGAGTGGCTTTTCTCAAGCTGGGTGAGGGTA 4718  
 |||:||||| |||:||||| |||:|||||  
 475 LysTrpLeuGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491  
 4719 T .....ACCTATAATGCT ..... 4731  
 |||:||||| |||:||||| |||:|||||  
 491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspVala 508  
 4732 .....GGTAGCCATGTTTCATCA 4749  
 |||:||||| |||:||||| |||:|||||  
 508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524  
 4750 GGTGGCTCGTCCCTAGAGGATATCTTAACGGCAATTAACTACCCAAG 4799  
 |||:||||| |||:||||| |||:|||||  
 525 Gly .....ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539  
 4800 CGTCTTCGACCACTGTTTAGTCAGAGAAGTACAGATGGGTGGAGCGCTAC 4849  
 |||:||||| |||:||||| |||:|||||  
 539 a .....GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552  
 4850 CATTCACCGCTGATGACGAGGAGTGTCTATCCATCAGATAACCTCATCTTG 4899  
 |||:||||| |||:||||| |||:|||||  
 553 .....LysValSerGlnValAspGlyArgIleAsp 562  
 4900 ACGGTCAATCTCGGGGAAGCAACTTCTCGACTAAGTGCAGGAAGGG 4949  
 |||:||||| |||:||||| |||:|||||  
 563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578  
 4950 TGGTAAGTCTAGTGTATACAGCTAGTTCGGGTGACTATTTCTTATGC 4999  
 |||:||||| |||:||||| |||:|||||  
 578 pGlyAlaValLeuGluThrAsn ..... 585  
 5000 CTTCGGCTTTTCAAGGACGCACCTTGCAT .....TCAGTAAACTCC 5040  
 |||:||||| |||:||||| |||:|||||  
 586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600  
 5041 ATCGACGAAGGGCGATCACTGTGACCTTCAGGCGCAACTCGGCGC ..... 5085  
 |||:||||| |||:||||| |||:|||||  
 601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617  
 5086 .....GTCCTTGGTGTAGGACGAGTGTTCGATTAGCCG 5119  
 |||:||||| |||:||||| |||:|||||  
 617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634  
 5120 GCGCGCTGCTGGATGAGAAGTACCAGGTGTCCA ..... 5154  
 |||:||||| |||:||||| |||:|||||  
 634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650  
 5155 .....AACACGACCAACACAG ..... 5169  
 |||:||||| |||:||||| |||:|||||  
 651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667

5170 .....AGCCAAGGTGCTACAGAACAAATCACACCAAAATCGG 5206  
 |||:||||| |||:||||| |||:|||||  
 667 nPheTrpPheHisProGluGlyLeuGlyProPheAlaProPheSerP 684  
 5207 GGGCAAGGCTCTATCTGAGGGAAGTGTAGGAAAGTCAAGGGAGGTCG 5256  
 |||:||||| |||:||||| |||:|||||  
 684 roGlyHisValTrpGlu.....SerAlaAsnPropheCysGlyGluSer 698  
 5257 ACATAC .....TCGATATGGTGGCAACAGATTACGTTAGGAAGTGTGA 5300  
 |||:||||| |||:||||| |||:|||||  
 699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715  
 5301 GTGG .....CTCAGGGCTG 5314  
 |||:||||| |||:||||| |||:|||||  
 715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732  
 5315 ATAATCCAGTGTGCTCTT .....GAACCTGACTACAC 5349  
 |||:||||| |||:||||| |||:|||||  
 732 laThrProThrProAlaAlaProLeuProProProAlaProAspProSer 748  
 5350 CCAATGACATTGGAAGTGGTAAACCGGACCTCTGAAGATGCCGCTG 5399  
 |||:||||| |||:||||| |||:|||||  
 749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764  
 5400 GGACTACTTGAAGTATCTGGCTATAGCATTGACAGGACATACAGGCGCT 5449  
 |||:||||| |||:||||| |||:|||||  
 764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781  
 5450 TG .....CTTATGGCTAGAAATATTCGCTC 5475  
 |||:||||| |||:||||| |||:|||||  
 781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797  
 5476 ACTACCGCGGAAGGTGTTCTGAAAGTACCTAAT ..... 5508  
 |||:||||| |||:||||| |||:|||||  
 798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814  
 5509 .....CAAGTTTATGAATCACTACCG ..... 5529  
 |||:||||| |||:||||| |||:|||||  
 814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831  
 5530 .....GGCTTTCACGTTTACAAAGTCGGCACA 5556  
 |||:||||| |||:||||| |||:|||||  
 831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847  
 5557 GATCTCATTTTTCATCAACAGAGCGGTGGTGTGTGAGAGACCTACC 5606  
 |||:||||| |||:||||| |||:|||||  
 848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864  
 5607 G ..... 5607  
 |||:||||| |||:||||| |||:|||||  
 864 oLysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881  
 5608 .....TAGTACTCATAGCTGAAAAGGTATCTTTACCAAG .....GGC 5646  
 |||:||||| |||:||||| |||:|||||  
 881 laAlaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896  
 5647 AAAGATGTCGACGCGGTGGTA .....GCTTTGGCGCAATCTCTT 5687  
 |||:||||| |||:||||| |||:|||||  
 897 ProSerPheAspAlaTyrGluArgAsnHisArgProGlyAspGluLeuTy 913  
 5688 CGTA .....TCGC 5695  
 |||:||||| |||:||||| |||:|||||  
 913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysP 930  
 5696 ACGATATACTGGTTTCCACCAT .....GCCATT 5724  
 |||:||||| |||:||||| |||:|||||  
 930 roThrLeuThrIleThrGluAspValAlaAlaArgThrAlaAsnLeuAlaIle 946  
 5725 AATTGTAGGTGCATGAAAGTCGCTGCATGC ..... 5757  
 |||:||||| |||:||||| |||:|||||  
 947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963  
 5758 .....GGCATGCTGGCGAATCGTTTAAGTCTCTCGAATATAAGTGCT 5800

963 lThrProGlyValVal.....GlnTyrGlnPhe 973  
5801 ATAATGCTCCCGAGGTGGGTAAAGACGACGAGTGTAGTGGACGAATTC 5850  
973 hrAlaGlyValProGlySerGlyLysSerArgSerIle..... 985  
5851 GTTAAGTACCCCAATAGCAGACCCACCAATTAGCGTAAATGTGGAGTTC 5900  
986 .....ThrGlnAlaAspValAspValValValProth 997  
5901 TCAGGACATAAATATGCGCGTGAAGAAGACAGATCCGAATTTGGAGGTC 5950  
997 rArgGluLeuArgAsnAlaTrpArgArg.....Glyp 1009  
5951 TCAACAGTCTACCCAGTAACTACAGGTGTAACTTTATCGTCAGG 6000  
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr..... 1021  
6001 GGAATGTATAAAGGTTTTGGTGGATGAGTGCACATGATGCATCAAGG 6050  
1022 ...GlnGlyArgArgValValIleAspGluAlaProSerLeuProProHi 1037  
6051 CTTACTACAACTAGGCGTCTTCGCAACCGCGCGTCGGAAGCCTCTTTT 6100  
1037 sLeuLeuLeuLeuHisMet.....GlnArgAlaAlaThrValHisLeuL 1052  
6101 TTGGAGACATAACATCAGATACCATTAACACAGGGAGAGGTGTTAGG 6150  
1052 euGlyAspProAsnGlnIleProAlaIleAspPheGluHis..... 1065  
6151 ATCGAATTGTGCTGTTTTTGTTCCTCAAG...AAGGAAAGCGTTGTATACAC 6197  
1066 .....AlaGlyLeuValProAlaIleArgProAspLeuAlaProth 1079  
6198 TTCT.....AAATCTACAGTGTCCGTAGATGTTTCTACTTGT 6238  
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095  
6239 TGTCTCAATGACGTGAAGGGAACGGAAAGTGTACCTGCAAAAGGTC 6288  
1096 .....IleArgGly.....AlaTyrProMetIleGln 1104  
6289 GTTAGCGTAAGGACAAACAGTAGTAGTACGCTG.....TCCAAAG 6332  
1105 ThrThrSerArg.....ValLeuArgSerLeuPheTrpGlyClnPr 1118  
6333 GCAATTGCAACCACTGATGACGTAGCTGAAATAAACGCTGACGTACT 6382  
1118 oAlaValGlyGlnLys.....L 1124  
6383 TGTGATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTGTTGAAGGA 6432  
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer..... 1137  
6433 AAAGGAAAAGAACACCACTGATGACGTGATGAAGCACAGGGGAAAAAC 6482  
1138 .....ValThrValHisGluAlaGlnGlyAlaTh 1147  
6483 ATTACGTATGCTGATTTAGGACGAAGAAAGCCGATGATCCCTAT 6532  
1147 rTyrThrGluThrThrIleAlaThrAlaAspAlaArgGlyLeuIleG 1164  
6533 TCACATAAACACCGCATATACTTGTGTTTTCGAGACACACACGCTCA 6582  
1164 InSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180  
6583 CTGGTTTATGCGCTCTGAGCTCAAAGTTGGACGATAAGTCGGCACATA 6632  
1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly.... 1194  
6633 TATTTAGCGACGCGTCACTCAATCATAGTATCCGACCGCTTGTGCTTCACT 6682  
.....

1195 .....IleSerAspAlaIleValAsnAsn 1203  
6683 TGCCTCCCGGTGGT 6696  
1203 hePheLeuAlaGly 1207  
seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-325-267A-2  
seq\_documentation\_block:  
; Sequence 2, Application US/08325267A  
; Patent No. 5585271  
; GENERAL INFORMATION:  
; APPLICANT: WATARI, JUNJI  
; APPLICANT: TAKATA, YOSHIHIRO  
; APPLICANT: OGAWA, MASAHIRO  
; APPLICANT: PENTTILA, MERJA  
; APPLICANT: ONNELA, MAIJA-LEENA  
; APPLICANT: KERANEN, Sirkka  
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325.267A  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP PCT/JP94/00290  
; FILING DATE: 24-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 38871/1993  
; FILING DATE: 26-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1537 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-325-267A-2  
alignment\_scores:  
Quality: 178.00 Length: 1508  
Ratio: 0.287 Gaps: 67  
Percent Similarity: 41.180 Percent Identity: 19.098  
alignment\_block:  
US-09-301-906-3/rev x US-08-325-267A-2 ..  
Align seg 1/1 to: US-08-325-267A-2 from: 1 to: 1537  
5504 GTACTTTTCAGACACCTTCGGGGTA.....GTGACGCAATATT 5464  
118 GlyPheTyrThrThrProThrAsnValThrLeuGluMetThrGlyTyrPh 144  
.....



5463 TCATGCCATAAGCAACGCCCTGTATGTCTCTCAATGCCTATAGCCAGAT 5414  
||||| .....:|||||  
144 eueProGlnThrGlySerTyThr..... 133

5413 ACTTCAGTACTCCACGAGCCATCTTCAGAGTC..... 5379  
|||||.....:|||||  
154 ..PheLysPheAlaThrValAspAspSerAlaIleLeuSerValGlyGly 169

5378 .....:CCGGTTTAAACCACTTC 5362  
|||||.....:|||||  
170 AlaThrAlaPheAsnCysCysAlaGlnGlnProPheIleThrSerThr 186  
|||||.....:|||||

5361 AAATGTCATTGGGTGTAGTCAGGTTCAAGAGCCATCACTGGATTATCAG 5312  
|||||.....:|||||  
186 rAsnPhe.....ThrIleAspGlyIleLysp 195

5311 CCCTGAGCCACTCA..... 5298  
:|||  
195 roTpgLysSerLeuProProAsnIleGluGlyThrValTyMetTy 211

5297 .....CACTTCCTAACCTAATCTTGTTCGCACCATATCAGAT 5260  
:|||.....:|||||  
212 AlaglyTyTyTyProMetLysValValTySerAsnAlaValSerTr 228

5259 TGTGCACCTCCCTTGACTTCCTACCACATCCCTCAGATAGAGCCTTGC 5210  
:|||||.....:|||||  
228 pGlyThrLeuProIle...SerValThrLeuProAspGlyThrThrValS 244

5209 CCCCAGATTGGTGTGATTCTTGTGTAGCACCTTGGCTCTGTGGTTGC 5160  
|||||  
244 erAspAspPheGlu..... 248

5159 TCGTTTGAACACCTGGTAGCTTCTCATCCGACAGCGCCGGCTAACTG 5110  
||| .....:|||||  
249 .....GlyTyrValTySerPheAspAspLeuSerGlnSerAsnCy 263

5109 CAACATCCTGCTACACCAAGAGCGCGAGTTGCCCTCAACGTCGAAC 5060  
:|||.....:|||||  
263 sThrVal...ProAspProSerAsnTyAlaValSerThrThrThrThr 279

5059 TGATGCGCCCTTCGTGATGGAGTTTACTGAATGCAAGTCGCGCTTTGA 5010  
:|||||.....:|||||  
279 hrThrGluProTrpThrGlyThrPheThrSerThr..... 290

5009 AAACCGCAAGCATAGAAAATAGTCAACCGAAGCTACGTTTATGACCAT 4960  
:|||||.....:SerThrGluMetThrThrValThrG 299

291 .....SerThrGluMetThrThrValThrG 299

4959 GACCTTA...CCACCCCTTCGTGACTTATGTCAGAGAGTTGGCTTCCCA 4913  
||| ||| .....:|||||  
299 yThrAsnGlyValProThrAspGluThrValIleValIleArgThrPro 316

4912 CGAGATTGACCGTCAAGATAGGTTATCTCATGATAGCACTCCTCGTCA 4863  
|| .....:|||||  
316 hrThrAlaSerThrIleIle.....ThrThrThr 325

4862 TCACGGTGGAAATGATAGCCCTCCACCACATCTTGACTTTCGACTAAACA 4813  
GluProTrpProAsnSerThr..... 331

4812 GTGGTCAAGACGCTTGGGTACTTAAATTCGGTTAAGATATCCTTAGGG 4763  
:|||||.....:|||||  
332 ..PheThrSerThrSerThrGluLeuThrThrValThrGlyThrAsnGlyV 348

4762 CACGAGGCCACCTTGATGAACATGGCTACCCACCATATAGTATACCCC 4713  
:|||||.....:|||||  
348 alArg.....ThrAspGluThrIleIle.....ValIleArg 358

4712 TCACCCAGCTTTGAGAAAAGACGGCACTCCTCCCTTCAAGTTATCGAC 4663  
:|||||.....:|||||  
359 ThrProThr.....ThrAlaThrThrAlaIleThrThrThrG 371

4662 AAAGGTAATCGCTTTGAGAACCTGGTGGCGATAGTCGCTTCTGTTGTACA 4511  
 371 uProPrAsn... SerThrPheT 378  
 4612 CATCCAGCTGCACACCGCAGGAACACAGCTTCTAAACATGCAAAAGCTGTTTA 4563  
 378 hrSerThrSerThrGluLeuThrValThrGlyThrAsnGly...Leu 393  
 4562 CCGGCCCTGGCTTCATGATGATCGTCAAACTCCTCTCTTTGGCGTGC... 4518  
 394 ProThrAspGluThrIleIleValIleArgThrProThrThrAlaThrTh 410  
 4517 ...TCGACTTGTA 4508  
 410 rAlaMetThrThrThrGlnProTPrAsnAspThrPheThrSerThrSerT 427  
 4507 CCGTGTCTTCACGCCACATGAGGACGGCTTGCC... 4470  
 427 hrGluMetThrThrValThrGlyThrAsnGlyLeuProThrAspGluThr 443  
 4469 TTGTCAATTATCTTCACACCACGAGTCGTAAACATGTTGGGCG... 4428  
 444 IleIleValIleArgThrProThrThrAlaThrThrAlaMetThrThrTh 460  
 4427 ...GTTTCTCATCGAGC...ATTGGCGTGGCGCCCG 4397  
 460 rGlnProTPrAspAspThrPheThrSerThrSerThrGluMetThrThrV 477  
 4396 AAACAGTCGCCAATGGT... 4380  
 477 alThrGlyThrAsnGlyLeuProThrAspGluThrIleIleValIleArg 493  
 4379 ...ACTTCTTGTCTTTCACCTTCTTCGAGCATGATCTCGAGCTGGGT 4336  
 494 ThrProThrThrAlaThrThrAlaMet...ThrThrThrGlnProTPrAs 509  
 4335 GACCTTGGGAATATCAGCGCTGTGGGAACTCACAAGACTGGACTGACTG 4286  
 509 nAspThrPheThrSerThrSerThrGluMetThrThrValThrGlyThrA 526  
 4285 GCNCCCTGTGCGCTTCGTTGACCGCGGTTGCTGTGTACCATCTGCA 4236  
 526 sGlyLeuProThrAspGluThrIleIleValIleArgThrProThrThr 542  
 4235 GCNCTT...CCCTCTCTCTTG 4219  
 543 AlaThrThrAlaMetThrThrThrGlnProTPrAsnAspThrPheThrSe 559  
 4218 TGTGGCGCTCACCGACGAGCTTCTTGGGAGCAACCTTTGGCGTCC... 4170  
 559 rThrSerThrGluMetThrThrValThrGlyThrAsnGlyLeuProThrA 576  
 4169 ...ACCTCAACCTCACGAGCGCTT 4149  
 576 spGluThrIleIleValIleArgThrProThrThrAlaThrThrAlaIle 592  
 4148 TCCCCACTAGCATGTTGGTGAAGACATPCGTCGACACCTTTTTCACACT 4099  
 593 ThrThrThrGluProTPrAsnSerThrPheThrSerThrSerThrGluLe 609  
 4098 TTCACAAAAGCCATGACGTGACGGGTACCAAGACCCGAC... 4059  
 609 uThrThrValThrGlyThrAsnGlyLeuProThrAspGluThrIleIleV 626  
 4058 ...GGATCACCTGTCTTTCGCCACAGCTGCA... 4029  
 626 alIleArgThrProThrThrAlaThrThrAlaMetThrThrThrGlnPro 642  
 4028 ...GACGGGCAACTCTTCTCTCGACATAAGTGATTGCACGCTTCG 3985  
 643 TrpAsnAspThrPheThrSerThrSerThr...GluMetThrThrValTh 658  
 3984 AGGTGAATGGGTCTTCTTCTCACCAAGTAACTTCCCGTGGTGCTAATCC 3935





```

4561 GGTAACAGCTTTGCGATGTTTGAAGCTGT.....TCCTGCGG 4598
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTyrLeuSerAlaG1 441
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4599 TGTGCAAGCTGGAT.....GTGTACAACGAAGCG..... 4626
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4627 .....ACTATCGCCACAGGCTTCTCAAAGCGCATTTACCTTTGTCGAT 4668
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
458 isCysArgThrAlaIleArgLysAlaLeuSerLysPheCysCysPheMet 474
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4669 AACTTGAAAGGAGGAGTGGCTCTTTTCTCAAAGCTGGTGAGGG... 4716
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyAl 491
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4717 .....TATACCTATAATGGT..... 4731
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
491 aValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4732 .....GGTAGCCATGTTTCATCA 4749
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4750 GGGTGCCCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAA 4799
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
525 Gly.....ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4800 CGTCTTCGACCACTGTTTAGTCAGAACTACAAGATGGGTGGAGGCGTAC 4849
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
539 a.....GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4850 CATTCACCGCTGATGACGAGAGTGCTATCCATCAGATAACCCCTATCTTG 4899
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
553 .....LysValSerGlnValAspGlyArgIleAsp 562
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4900 ACGGTCAAATCTCGTGGGAAGCAAACTTCTCGAATAAGTCAGGAAGG 4949
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4950 TGGTAGGTCATGGTCATAAACCTAGCTTCGGGTGACTATTTTCTATGCC 4999
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
578 pGlyAlaValLeuGluThrAsn..... 585
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5000 CTTGGCGTTTTCAAAGGACCGCATTTGCGAT.....TCAGTAAACTCC 5040
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5041 ATCGACGAAGGCGCATGATGTTTGAGTTTCAGGCAACTCGGCGC..... 5085
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5086 .....GTCTTTGGTGTAGGCAGGATGTTGCGATTTAGCCG 5119
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5120 GCGGCGTGTGCGATGAGAAGTCACCAAGT..... 5148
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5149 .....GTTCCAAA 5156
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgHisSerLeuIleGlyAs 667
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5157 CCAGCAACACAGACCAAGGTGCTACCAGACAATCACACCAAAATCGG 5206
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
667 nLeuTrpPheHisProGluGlyLeuIleGlyLeuPheAlaProPheSerP 684
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5207 GGGCAAGGCTCTATCTAGGAGGAGTGGTGGAGAGTCAAGGGAGGTCG 5256
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
684 roGlyHisValTrpGlu.....SerAlaAsnPropheCysGlyGluSer 698
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5257 ACATAC.....TCGATATGGTGGCAACAAGATTACGTTAGGAAGTGTGA 5300
| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
699 ThrLeuTyrThrArgThrTyrSerGluValAspAlaValSerSerProAl 715
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5301 GTGG.....CTCAGGCGTCG 5314
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
715 aArgProAspLeuGlyPheMetSerGluProSerIleProSerArgAlaA 732
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5315 ATAATCCAGTGTGGCTCTT.....GAACTGACTACACC 5349
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
732 laThrProThrLeuAlaAlaProLeuProProProAlaProAspProSer 748
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5350 CCAATGACATTTGAAGTGTAAACCCGGGACCTCTGAAGATGCCGTCGT 5399
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
749 Pro.....ProProSerAlaProAlaLeuAl 757
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5400 GGAGTACTTTGAAGTATCTGCTATAGGCATTGAGAGGACATACAGGCGT 5449
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
757 aGluProAlaSerGlyAlaThrAlaGlyAlaProAlaIleThrHisGlnT 774
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5450 TGGTTATGGCTAGAAATATTGCGCTACCTACCGCCGAAGGTGTTCTGAAA 5499
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
774 hrAlaArgHisArgLeuLeuPheThrTyrProAspGlySerLysVal 790
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5500 GTACCTAATCAAGTTTATGAATCACTA..... 5526
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
791 PheAlaGlySerLeuPheGluSerThrCysThrTyrProAspGlySerLys 807
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5527 .....CCGGGC.....TTTCACG 5539
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
807 rAsnValAspHisArgProGlyGlyGlyLeuCysHisAlaPheTyrGlnA 824
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5540 TTTCAAGTCGGGCACAGATCTCATTTTTCATTTCAACACAAGACGCG... 5586
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
824 rgTyrProAlaSerPheAspAlaAlaSerPheValMetArgaspGlyAla 840
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5587 .....TTGCGTGTGAGAGACCTACCGTACGTACTACTACCTGA 5624
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
841 AlaAlaTyrThrLeuThrProArgProIleIleHisAlaValAlaProAs 857
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5625 AAAAGGTATCTTTACCAAGGGCAAGATGTCGACGCGGTGGTAGCTTTGG 5674
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
857 pTyrArgLeuGluHisAsnProLysArgLeuGluAla..... 869
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5675 GGCACAATCTGTTGCTATGCGACGATATCTGTTTTCACGATGCCATT 5724
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 .....AlaTyrArgGluThrCys 875
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5725 AATTTGATAGTGCACCTGAAAGTCGCTCGATGCGGCATG..... 5763
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
876 SerArgLeuGlyThrAlaAlaTyrProLeuLeuGlyThrGlyIleTyrG1 892
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5764 .....GTGGGGAATCGTTTAAGTCTTCAATATACTGCTATAATC 5806
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
892 nValProIleGlyProSerPheAspAlaTrpGlu.....ArgAsnH 906
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5807 CTCCCCCAGGTGGC..... 5820
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
906 isArgProGlyAspGluLeuTyrLeuProGluLeuAlaAlaArgTrpPhe 922
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5821 .....GGTAAGACGACGCTTAGTGACGAATTCGT 5852
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
923 GluAlaAsnArgProThrArgProThrLeuThrIleThrGluAspValAl 939
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5853 TAAGTCACCAATAGCAGACCCACCATACGGCTAATGTGGGAAGTTCG 5902
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
939 aArg.....ThrAlaAsnLeuAlaIleGluLeuAspSerAlat 952
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5903 AGGACATAAATATGCGGTGAAG..... 5925
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
952 hrAspValGlyArgAlaCysAlaGlyCysArgValThrProGlyValVal 968
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5926 .....AAGAGAGATCCGAATTTGGAAAGGTCTCAACAGTGTACCACAGT 5969
| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

969 GlnTyrGlnPheThrAlaGlyValProCglySerGlyLysSerArgSerIle 985  
5970 TAACTCCAGGGTGTAACATTTATCGTC..... 5997  
      : : : : : | : : : :  
985 eThrGlnAlaAspValAspValValValProThrArgGluLeuArgA 1002  
5998 .....AGGGCAATGTAT..... 6009  
      | : : : : :  
1002 snALeTrpArgArgGlyPheLaLaPheThrProHisThrAlaAla 1018  
6010 .....AAAAGGTTTTGGTGATGAGTCGCATCATGTCGA 6044  
      : : : : : | : : : : | : : : :  
1019 ArgValThrGlnGlyArgValValIleAspGluAlaProSerIeuPr 1035  
6045 TCAAGCGTTACTACAACCTAGGCGCTTCGCAACGGCGCGTCGGAAGGCC 6094  
      | : : : : | : : : : | : : : :  
1035 oProHisLeuLeuLeuLeuHisMet.....GlnArgAlaAlaThrValH 1050  
6095 TCCTTTTTTCGGAGACATAAATCAGATACCATTATCAACACGGGACAAGCTG 6144  
      : : : : : | : : : : | : : : : | : : : :  
1050 IsLeuLeuGlyAspProasnGlnIleProAlaIleAspPheGluHis... 1065  
6145 TTTAGGATGGATTGCTGCTTTTGTGCCAAG...AAGAAAGCGTTGT 6191  
      | : : : : | : : : : | : : : :  
1066 .....AlaGlyLeuValProAlaIleArgProAspLeuGl 1077  
6192 ATACACTTCT.....AAATCGTACAGTGTCCGTTAGATGTTGCT 6232  
      | : : : : | : : : : | : : : : | : : : :  
1077 yProThrSerTrpTrpHisValThrHisArgTrpProAlaAspValCysG 1094  
6233 ACTCTGTGTCCTCAATGACCGTAAAGGGAACGGAAAAAGTGTACCCGTAA 6282  
      | : : : : | : : : : | : : : : | : : : :  
1094 luLeu.....IleargGly.....AlaTyrPromet 1102  
6283 AAGTGTGTTAGCGTPAAGACAACACAGTAGTAGATCGCTG.....TC 6326  
      : : : : : | : : : : | : : : : | : : : :  
1103 IleGlnThrThrSerArg.....ValLeuArgSerLeuPheTrpGl 1116  
6327 CAAAGGCCAATTGGAACCACTGATGACGTAGCTAGCTGAATTAACCGCTCACG 6376  
      : : : : : | : : : : | : : : : | : : : :  
1116 yGluProAlaValGlyGlnLys..... 1123  
6377 TGCTACTTGTGCATGACCCAGTTGGAGAAGTCGCATATCAAGAGGTCTGTTG 6426  
      | : : : : | : : : : | : : : : | : : : :  
1124 ....LeuValPheThrGlnAlaAlaLysProAlaAsnProGlySer... 1137  
6427 AAGGAAAGGAAGAACAACACAGTGTGATGACATGAAGCAGCAGGG 6476  
      : : : : : | : : : : | : : : : | : : : :  
1138 .....ValThrValHisGluAlaGlnGl 1145  
6477 NAAAAATTCAGTGATGTGTATTGTTTAGGACGAAGAAGCCGATGACT 6526  
      | : : : : | : : : : | : : : : | : : : :  
1145 yAlaThrTyThrGluThrThrIleIleAlaThrAlaAspAlaArgGlyL 1162  
6527 CCTATTCTCAATAAACCCGATATCTGTTGTTTTCGAGACACACA 6576  
      : : : : : | : : : : | : : : : | : : : :  
1162 euIleclnSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThr 1178  
6577 CGCTCACTGGTTTATGCGCTCTCGAGCTCAAGTTGGACCATTAAGGTCGG 6626  
      : : : : : | : : : : | : : : : | : : : :  
1179 GluLysCysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGl 1194  
6627 CATCATATATTAGCGACGCGTCACCTCAATCATGATTCCTCGAGCTTTGCTTC 6676  
      | : : : : | : : : : | : : : : | : : : :  
1194 y.....IleSerAspAlaIleVala 1201  
6677 ACACGTCGCCCGCGCTGGT 6696  
      : : : : : | : : : : | : : : : | : : : :  
1201 snAsnPheLeuAlaGly 1207

seq\_name: /cqn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-616-844-40

seq\_documentation\_block:

```

: Sequence 40, Application US/08616844
: Patent No. 5849578
: GENERAL INFORMATION:
: APPLICANT: FALB, DEAN A.
: TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
: NUMBER OF SEQUENCES: 54
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/616.844
: FILING DATE: 15-MAR-1996
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/599,654
: FILING DATE: 09-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/485,573
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/386,844
: FILING DATE: 10-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CORUZZI, LAURA A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-053
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1481 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-616-844-40

```

```

alignment_scores:
  Quality: 161.50      Length: 1318
  Ratio: 0.273         Gaps: 67
  Percent Similarity: 44.917  Percent Identity: 20.865

alignment_block:
  US-09-301-906-3/rev x US-08-616-844-40 ..
  Align seg 1/1 to: US-08-616-844-40 from: 1 to: 1481

3852 ATGCTGAGTCC.....AGATGGACAGACCTTCGTATAGTCAA 3812
   ||| |||||
1 MetaSerProArgAlaSerArgTrpProProLeuLeuLeuLe 17

3811 AGTCCCGGTAACTACCTAAGCGTCTCATCGTACACATTCGTCCTCCTACC 3762
   :|||
   :|||
17 uleuPro.....LeuLeuLeuLeuPro..... 25

3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTTCGTTAAGCAAGCCCTTAT 3712
   |||||
26 .....AlaAlaProGly 29

3711 GATCAAAAGATACCCTTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662

```

```
30 ThrArg.....|||
3661 TCCTCGCCAAACAACCTCGGAAACAGC...GCACCAATTATTGGAGAAG 3615
34 oProSerProIaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAla 51
3614 CTTCCAAACTTCCAAGTATTTCGGCGTAGCCCTCGCCCAAGCAATGTCCC 3565
51 LyLeuGluLeuGlnLeuGluArg.....ArgProGluArg.GluPr 64
3564 TAAACCAAGAAACCAATTAAAGTCGCTGGAGAGCATGTGCCCAATGC 3515
64 oProProThrProProArgGluArgGlyProAla...ThrProGlyP 80
3514 CTAACATACTGCAGCCGAA..... 3495
80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96
3494 .....ATGGCGTATGAGGTAGAAAA 3475
97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113
3474 CGATACAGCGCTAAGGGACTTGACGTTTACGTACGTCGCCCTAGTGA 3425
113 oGluSerAsnThrGlu.....AlaHisValGluAsnIleT 125
3424 CGGTACCGCGGACGAAACCCCTCTTCGCAGCACTAAAGATCACCGGAT 3375
125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141
3374 ACTGTCTAGAGGTTAAACCGCTAAACACCACCGCGCTGTGAACTTAA 3325
142 ValMetValGlnThrSerGlyLysSer..... 150
3324 CGAGCTTAGAGTTTCGTAACTACTGCTCCCGGTACCTTCTAGCGTTAGCCG 3275
150 ..... 150
3274 CAATTCGGCGCTATACTTCCATGCAATTGGGCGCGCACCCAAAGACGC 3225
151 .....HisAlaAlaSerAspAlaProGluAsnLeu 160
3224 ACGAATATCCCGATAGCAACCTTATTCGCAATACGAAAGTGGTGAAGTC 3175
161 .....ThrLeuLeuAlaGluThrAlaAspAlaArgGI 171
3174 ATTTAAGGAACCTTCACTCCAAATTTCTTTCGTTCCAGCTACGACCTTGC 3125
171 yArgSerGlySerSerSerArgThrAsnPhe..... 181
3124 TCACAGCTAGCGCGAGCTCTCAGTATGCCACCGGTGAGTTT..... 3081
182 .....ThrIleLeuProValGlyTyrSerLeuGlu 191
3080 ATAGCATCTCTGTGCGCACGAGCATCAAGCCTTTTCATCT...TCAGT 3034
192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208
3033 TAAACCGCGGTCTAGAAATAGTGAACCTGTGGAAAGAACGAGCTAATGC 2984
208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225
2983 CAGAAAGGCCCTTCGTCGAGCCCGACGAGCTACCAACCCCTTAGAAAA 2934
225 InThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241
2933 CTCCAA.....ACCACGTCGCTTT 2914
242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258
2913 CGAACCGCCCAAGAACCTGCCCTC..... 2889
:::|||||
```

```
258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275
2888 .....GTACAGTGG..... 2880
275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291
2879 ...ACGTTAGAAACGTTCGCCATTATGAAGCATATACAGAATCTTTCAGT 2833
292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu... 307
2832 AATTGGCAGGGTTTCCATCTTCTCTCTTCATTCGCTTCATGACCTTAA 2783
308 .....SerThrGluLysLeuAsnAsnSerThrGlyL 318
2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCGC.....AAGCGACTA 2739
318 euGlnSerSerValSerGlnThrLysThrMethHisValAlaThrVal 334
2738 TTCTTCAGTTTCTACCGAAGGTC.....TCCTTCGAAATATCAAAGGC 2695
335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGI 351
2694 ACCCTGAGCTCCAACTGCTGCTGAAGTCCAGAGTGCAAAATGGTAACAA 2645
351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368
2644 CAACGACCCAGGTACTTTACGGATGAACGCTCTGGGC..... 2607
368 hrSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384
2606 .....ATAGACATAAGCGT 2593
385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401
2592 CTCCAAAACGTTGAAGTTATCGTGCACCATTCGCTTAGCATAGACGAGA 2543
401 rGluAsnGluPheGlyLeu.....ThrSerL 410
2542 CTGCGTTGAAAAACAATCGCTCCTGCTGAAAGAAGAGCTTGAAGACC 2493
410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426
2492 CCGCGGAGACTCCCTCCTACTGGCTTTTATAAGTTATAGGAGCGAATACT 2443
427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu..... 440
2442 CTTGCGCCTATCCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393
441 .....ThrValSerArgSerValAlaPro. 448
2392 ACATATCATCAGGAGAAATATCAACCTTGCCTGG..... 2358
449 ..MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463
2357 ACAACGCGACTCCCTCTGA..... 2340
464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480
2339 .....ATGACATGGGTCTTTTGGCGACCTGACC..... 2313
480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497
2312 ..GCCCAACGCGATACTCAACGTTCTATTCTCGAATGTCCCTTAGTAGT 2265
497 lYSerHisThrAlaLeuGlyAspSerTyrSerGluSerSerSerThr 513
2264 TTGACGCGCTAGGAATAGATCCTGTCGACAAAGTCCGAAATCGAGGACTAT 2215
514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529
2214 TGTCTTAAAGGACACACTAGAACCA.....G 2189
529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSera 546
:::|||||
```

[illegible]

1366 .....CCTTTCCCGCGCTTCAGAACACATCGCA 1366  
770 rSerSerGlyProLeuProSerValSerGlnSerHis.H1 786  
1365 CCTATCTGAGCGTTTTCGCGAATCATTTCCGCTTGTGAGAAACATAT 1316  
786 sleu.....PheSerSerIleLeuProSerThrArgAlaSerVal. 799  
1315 CGCACAATTTTCGTCCAAATGGGAAGCAATGCGATTTGGTCCTTAAAT 1366  
800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815  
1265 ACAGTCTATTGGTGATCTACTGAAACGACCGATGAATAGACCTCTCTTG 1216  
816 ProSerProLeuProValSerLeuThrThrSer.....ThrSerAlaPr 830  
1215 ATTAAGGCGATTGGCGACGAGGCCAAGATCTCCCTCCCAACCGCTGCC 1166  
830 oleuSerValSerGlnThrThrLeuProGlnSerSerSerThrProVal 847  
1165 TATCGCCACGACGATACGGCTTTTGGGACATTTGGGAGATAGACATCT 1116  
847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863  
1115 CGCGCCTTCCACCCAAAGGTACGCTTTCATTGATGATGCGCATGTGTAG 1066  
864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880  
1065 TATGTTGTCGCCCGCAACCTTAATGACCTTTAACGATCTAAAG...CCAA 1019  
880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897  
1018 GGTGATCAAGAAACCTATCGAAAG..... 993  
897 erLeuValSerLeuProThrGlnSerThrLysAlaValThrThrAsnSer 913  
992 .....GAAGATGAGAAAGCAAGCGCTCTGTAAAT 964  
914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928  
963 TTCTCCCTCAGCCTCTCATTTGTTAAACCAACACACCCCTTTGGTGGTGC 914  
929 ....ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThr 944  
913 ATACAAACGCCCTTTCGGCGATCTCCACAACTCAATGCGCGCAAGGCA 864  
944 hrThr.....IleLeuLysThrSerGlnProLeuMetThr..... 955  
863 ACATCACACCCCTGGGCTGTATTACGATGAAGTCCTCGATC..... 822  
956 Thr.....ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970  
821 .GTACCCCTACCCGCGACAATAGGAGAAATTTAACAGCTCATTTTGA 773  
970 ofIleAlaValGlnThrThrAlaGlyLysGln.....LeuSerLeu 984  
772 CGTAAATAACCTCTCTCTACGAGGCTTCATCCAG.....ATTTTA 729  
984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyIleSer 1000  
728 GCCCAAGCAAGCGTGTCTTTACATCAATAGCAATTAAGGCGCTGTT 679  
1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017  
678 AAACAATTTCCCT.....TTCCGCAAGATGACAGCTCGCTCTCA.... 639  
1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyVal 1034  
638 .....TCACCCCCCTTACTTGAGAGAAAGGCCAACCGTCACG 600  
1034 hrAlaGluThrSerPro.....AlaSerArg.SerLeuGlyThr 1046  
599 GCGCGCTCTCTCCCAAGCACTGTTGTGGGACT.....GATTGCGCTG 557

1046 rSerProSerProGlnThrValValSerThrAlaGluAspLeuAlap 1063  
556 GTCCTCAGTCTCTCTCTTTTGGAGAGTGGTACGTGGGACGCAACAGCT 507  
1063 rOlySerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079  
506 GTTGGCTTAACAGCCCAATGTGACGCTTCCGCTTGGGCTTCTGC 462  
1080 LeuSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-599-654-40

seq\_documentation\_block:  
; Sequence 40, Application US/08599654  
; Patent No. 5882925  
; GENERAL INFORMATION:  
; APPLICANT: FALB, DEAN A  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,654  
; FILING DATE: 09-FEB-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,573  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/386,844  
; FILING DATE: 10-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CORUZZI, LAURA A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-041  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1481 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-599-654-40

alignment\_scores:  
Quality: 161.50 Length: 1318  
Ratio: 0.273 Gaps: 67  
Percent Similarity: 44.917 Percent Identity: 20.865

alignment\_block:

US-09-301-906-3/rev x US-08-599-654-40 ..

Align seg 1/1 to: US-08-599-654-40 from: 1 to: 1481

3852 ATGCTGAGTCC.....AGATGGACCAAGCCTTCGTCTATGCTCAA 3812  
||| ||||| |||||

1 MetAlaSerProArgAlaSerArgTrrProProLeuLeuLeuLe 17  
3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCTCTACC 3762  
17 uLeuPro.....LeuLeuLeuLeuProPro..... 25  
3761 GGTGGGACGTAAGCAAGTTGAGGTATGCTGCTTAAGCAAGCCCTTAT 3712  
26 .....AlaAlaProGly 29  
3711 GATCAAGATACCCCTAGCGTATATAGCGAGTACAAATGAGCTGACACTCT 3662  
30 ThrArg.....AspProPr 34  
3661 TCCTCGCCCAACAACCTCGGGAACAGC...GCACCAATATTGGAGAG 3615  
34 OProSerProAlaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51  
3614 CTTCCAAACTTCCAAGTATTTCGGCGTAGCCCTCGCCCAAGCAATGTCC 3565  
51 LylLeuGluLeuGlnLeuGluArg.....ArgProGluArg.GluPr 64  
3564 TAAACCAAGAAAACATTAAAGTCGCGTGGAGAGCATGTGCCCAATGC 3515  
64 OProProThrProProArgGluArgArgGlyProAla...ThrProGlyP 80  
3514 CTAACATACTGCAGCCGAA..... 3495  
80 rSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96  
3494 .....ATGGCGTATGAGCTAGCAAAA 3475  
97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113  
3474 CGATAGCAACGTTAAGGGACTTGAGCTTTAGCTAGCGTCGCGCTAGTGA 3425  
113 oGluSerAsnThrGlu.....AlaHisValGluAsnIleT 125  
3424 CGGTACCCCGCAGCAAAACCTCTTCGCGAGCACTAAAGATCCACCGGAT 3375  
125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141  
3374 ACTGCTAGAGGTTAAACCGCGCTAAACCAACCCGCGCTTGTGAACCTAA 3325  
142 ValMetValGlnThrSerGlyLysSer..... 150  
3324 CGAGCTTAGAGTTTCGTAACACTACTGCCCGGTACCTTCTAGCGTTAGCG 3275  
150 ..... 150  
3274 CAATTCGCGCCTATATTCCATGCAATTGGGCGCGCACCCCAAGAGCGC 3225  
151 .....HisAlaAlaSerAspAlaProGluAsnLeu 160  
3224 AGCAATATCCGATAGACACCTTATTCGCAATATCAAGTGGTGAAGTC 3175  
161 .....ThrLeuLeuAlaGluThrAlaAspAlaArgG 171  
3174 ATTTAAGGAACTTCACTCCAAATTTCTTTCGTTCCAGCTACGACCTTGC 3125  
171 yArgSerGlySerSerArgThrAsnPhe..... 181  
3124 TCACAGCTACGCGGACGTCTCCAGTATGCCACCGGTGAGTTT..... 3081  
182 .....ThrIleLeuProValGlyTyrSerLeuGlu 191  
3080 ATAGCATTTCTCTGCGCACAGAGCATCAAGCTTTTCATCT...TCAGT 3034  
192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208  
3033 TAAACCGCGGTCTAGATAGTGAACCTGTGGAAAAGACGAGCTAATGC 2984  
208 uHisLeuProSerSerSerGluPheAspGluArgIleAlaAlaPheG 225



[illegible]

2312 ..GCCCAACGGCATCTCAAAGGTTCTATTCTCGAATTCCTCCTATAGTG 2265  
::: ||||| ::::: ::|||::: :::  
497 lySerHisThrAlaLeuGlyAspArgSerTyrSerCluSerSerSerThr 513  
  
2264 TTGAGCCGGTAGGAATAGACTCTGFCGACAAGTCCGAATCGAGGACTAT 2215  
||||::: ||| :::: ||||::: ::::: :::  
514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529  
2214 TGCTTTAAAGGACACACTAGAACCA.....G 2189  
529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerAla 546  
2188 CTCACAGAGGTATGAACACCAGCGAGTTGGCCCAACGACGCGGACCAAC 2139  
laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561  
2138 GACCTATACGTAACCTCAGT.....TCCAGGTGACCTCTCGAGCCTAC 2096  
||| ::::: ::::: ||| ||| ||| ||  
562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578  
2095 ACATAGTGAGATGTTGATAACCAACACAGGTATCCCTCATACTCTAGCTTA 2046  
| ::::: ::::: |||||  
578 rLeuThrAsnThrSerMetSerThr..... 587  
2045 AACACGTTCCGGGACGAAGTCGTACATAGAANAAGGTGA.....AAA 2002  
588 .....SerGlyGluAlaGlySer 593  
  
2001 CCGCTTACGTT.....AGACAAGGAATGTTTCATCATCTCAC 1964  
||||| ::: |||||  
594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609  
1963 CACACTCACCAACATGTTAGCGCCAGGTAAATCACCTTCCCCTCTTCACCAAC 1914  
::: ||::: ||::: ||| |||  
610 AsnValThrAspAspMetGlyLeuVal.....SerAla 620  
  
1913 GT.....ATCAGAGTCTTCT 1900  
|| |:::|  
620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636  
1899 CAAGTAACCTTACCCTTACCCTACACACATCCACGGGGACAT.... 1855  
||||::: ||::: ||::: ::::: |||  
637 GluValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653  
1854 .....AAGACTATGTCGAACACAGCGCACCTTCTCTTA 1821  
653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669  
1820 TCCATAGCCTCTGCTACTGCTGGGGGATATA.....TCGTACACATC 1777  
||||::: ::::: |||||  
670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrse 686  
1776 CACCATAAACCCGCATCGGCTCTAGATTCGAGTTCTGCGATTCCTCGAC 1727  
| ||||| ::::: ::::: |||||  
686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr..... 698  
1726 TACAGTAAGACACAGACTTTCAAGCCGCTTCAGTTAGAGGTCA...CTG 1680  
699 .....SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711  
1679 GACACGTACGAA.....TCTCCCCCAGCTGGGAAGAAGAGGAT 1639  
::: ||::: ||::: |||||  
712 SerSerThrAspGlyGluTyrAlaGlnProSerThrGlu..... 724  
1638 CTGATTGATTCTCGGCTTACATCTTTAACGTCCTAGGACTGGATTGCATA 1589  
||| ::::: ||::: ||| |||  
725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr..... 737  
1588 CATGCAGTTTCACATGGCCACCTTTGACATGATACGTGAGCTCCCCCCA 1539  
|||  
738 .....ProThr 739  
1538 ATATCGAAGAAGCACACATTACCACACCTCTCTGGAAAAGATTCCAATTGA 1489



Quality: 161.50 Length: 1318  
Ratio: 0.273 Gaps: 67  
Percent Similarity: 44.917 Percent Identity: 20.865

## alignment\_block:

US-09-301-906-3/rev x US-08-944-868A-40 ..

Align seg 1/1 to: US-08-944-868A-40 from: 1 to: 1481

```
3852 ATGCTGAGTCCC.....AGATGACCAGCACCTTCCTCATAGTCAA 3812
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetAlaSerProArgAlaSerArgTrrProProLeuLeuLeuLeuLe 17
3811 AGTCCCGGTAAATACCTAAGCGTCTCGTACACATTCCTCCCTCTACC 3762
||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uLeuPro.....LeuLeuLeuLeuProPro..... 25
3761 GGTGGGAGCTAAGCAAGTTGAGGTATGGTGTCTTAAGCAAGCCCTTAT 3712
||||| ||||| ||||| ||||| ||||| ||||| |||||
26 .....AlaAlaProGly 29
3711 GATCAAGATACCCCTAGCGTAATAGGCAGTAACATGAGCTGACACTCT 3662
||| ||||| ||||| ||||| ||||| ||||| |||||
30 ThrArg.....AspProPr 34
3661 TCCTCGCCNAACACCTCGGACAGC...GCACCAATTATTGGAGAG 3615
||||| ||||| ||||| ||||| ||||| ||||| |||||
34 oProSerProAlaArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51
3614 CTCCCAAACTTCCAGTATTTCGGCTAGCCCTCGCCCAAGCAATGTCCC 3565
||| ||||| ||||| ||||| ||||| ||||| |||||
51 lLeuGluLeuGluLeuGluArg.....ArgProGluArg.GluPr 64
3564 TAAACCAAGAAAACATTAAAGTCCGCTGGAGAGCATGTGCCCAATGC 3515
||| ||||| ||||| ||||| ||||| ||||| |||||
64 oProProThrProProArgGluArgGlyProAla...ThrProGlyp 80
3514 CTAACATCTGCAGCCGGAA..... 3495
||||| ||||| ||||| ||||| ||||| ||||| |||||
80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96
3494 .....ATGGCGTATGAGTAGAGAAA 3475
||||| ||||| ||||| ||||| ||||| ||||| |||||
97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisrrPr 113
3474 CGATAGCAACGCTAAGGGACTTTGACGTTTACGTCGCGCCCTAGTGA 3425
||||| ||||| ||||| ||||| ||||| ||||| |||||
113 oGluSerAsnThrGlu.....AlaHisValGluAsnIleT 125
3424 CGTCAACCGCGAGCAAAACCCCTCTTCGCACGACTAAGATCCACGGAT 3375
||| ||||| ||||| ||||| ||||| ||||| |||||
125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerLysGluGly 141
3374 ACTGTGCTAGAGTTAAACCGCGCTAAACACCGCGCGCTTGTGAACCTAA 3325
||||| ||||| ||||| ||||| ||||| ||||| |||||
142 valMetValGlnThrSerGlyLysSer..... 150
3324 CGAGCTTAGAGTTTCGTAACCTACTGCCCGCGTACCTTCTAGCGTTAGCCG 3275
||||| ||||| ||||| ||||| ||||| ||||| |||||
150 ..... 150
3274 CAATTCGCGCCTATCTCCATGCAATTGGGCGCGCACCAAGAGGCC 3225
||||| ||||| ||||| ||||| ||||| ||||| |||||
151 .....HisAlaAlaSerAspAlaProGluAsnLeu 160
3224 ACGAATATCCGATAGCACTTATTCGCAATACGAAAGTGTGAAGTC 3175
||||| ||||| ||||| ||||| ||||| ||||| |||||
161 .....ThrLeuLeuAlaGluThrAlaAspAlaArgGly 171
3174 ATTTAAGAAACTTCACTCAAAATTTCTTCGTTCCAGCTACGACCTTGC 3125
||| ||||| ||||| ||||| ||||| ||||| |||||
171 yArgSerGlySerSerSerArgThrAsnPhe..... 181
```

```
3124 TCACAGCTACGCGCAGCTCTCCAGTATGCCACCGGTGAGTTT..... 3081
||| ||||| ||||| ||||| ||||| ||||| |||||
182 .....ThrIleLeuProValGlyTyrSerLeuGlu 191
3080 ATAGCATTTCTGTGCGCACAGCATCAAGCCTTTTCATCT...TCAGT 3034
||||| ||||| ||||| ||||| ||||| ||||| |||||
192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208
3033 TAAACCCGGTCTAGAAATAGTGAACCTGTGGAAAGAACGACCTAATGC 2984
||| ||||| ||||| ||||| ||||| ||||| |||||
208 uHisLeuProSerSerSerGluPheAspGluArgIleAlaAlaPheG 225
2983 CAGAAAGGCTTCGTGCGACGCCACAGAGCTACCAACCCCTTAGAACA 2934
||||| ||||| ||||| ||||| ||||| ||||| |||||
225 lnrThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241
2933 CTCGAA.....ACCACGGTCGCTTT 2914
||||| ||||| ||||| ||||| ||||| |||||
242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258
2913 CGAACCCCAAGAGACCTGCCCTC..... 2889
||||| ||||| ||||| ||||| ||||| |||||
258 pSerProSerPheLeuProAlaLeuGluMetGlyLeuThrThrProS 275
2888 .....GTACAGTGG..... 2880
||||| ||||| ||||| ||||| ||||| |||||
275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291
2879 ...ACGTTAGAAAGCTTCCCATTTATACAGATATACAGAACTTTCAGT 2833
||| ||||| ||||| ||||| ||||| ||||| |||||
292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu.. 307
2832 AATTGGCAGGGTTTCCATCTTCTTCATTCCTTCATGACCTTAA 2783
||| ||||| ||||| ||||| ||||| ||||| |||||
308 .....SerThrGluLysLeuAsnSerThrGlyL 318
2782 TTGNAATCTTCCACCATAGCCCTAGAGAGAGCGC.....AAGCGACTA 2739
||||| ||||| ||||| ||||| ||||| ||||| |||||
318 euGlnSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334
2738 TTCTTCAGTTTCTTACCGAAGGTC.....TCCTCGAAATATCAAGGC 2695
||| ||||| ||||| ||||| ||||| ||||| |||||
335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGl 351
2694 ACCCTGAGTCCAACTGTCTGAAGTCCCAAGTCCAAATGCTAAACA 2645
||||| ||||| ||||| ||||| ||||| ||||| |||||
351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThr 368
2644 CAACAGACCCAGGTACTTTACGGATGAACGCTCTGGGC..... 2607
||||| ||||| ||||| ||||| ||||| ||||| |||||
368 hrSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384
2606 .....ATAGACATAAGCGT 2593
||||| ||||| ||||| ||||| ||||| ||||| |||||
385 SerArgValThrGlyAsnProGlyAspGluGluPheLeuGluProSerTh 401
2592 CTCCAAACGTTGAAGTTATCGTCACCATTCCTTAGCATAGACGAGA 2543
||||| ||||| ||||| ||||| ||||| ||||| |||||
401 rGluAsnGluPheGlyLeu.....ThrSerL 410
2542 CTGCGTTCGAAACAATCGCTACGCTCTGAAGAAGAGAGCTTGAAGACC 2493
||||| ||||| ||||| ||||| ||||| ||||| |||||
410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426
2492 CCGCGAGACTCCCTCCTACCTGGCTTTTATAAGTTATAGGAGCGCAACT 2443
||||| ||||| ||||| ||||| ||||| ||||| |||||
427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu..... 440
2442 CTTTCGCCCTTCTTAATCGCTGAGCATAACACGACGACAACTAAACCCC 2393
||||| ||||| ||||| ||||| ||||| ||||| |||||
441 .....ThrValSerArgSerValAlaPro. 448
2392 ACATATCATCAGGAGAAATATCAACCTTGTGTGG..... 2358
```





150 ..... 150  
3274 CAATTCCGGCCTATATCTCCATGCAATTGGCGCGCACCCCAAGACGCC 3225  
151 ..... HisAlaAlaSerAspAlaProGluAsnLeu 160  
3224 AGGAATATCCGATAAGCACCTTTATCCGCATATACGAAAGTGGTGAAGTC 3175  
161 ..... ThrLeuLeuAlaGluThrAlaAspAlaArgG1 171  
3174 ATTTAAGGAACTTCACTCACAAATCTTTCTTCCTCCAGCTAGACCTTGC 3125  
171 yArgSerGlySerSerArgThrAsnPhe ..... 181  
3124 TCACAGCTACGGCGCAGCGTCTCCAGTATGCCACCGGTGAGTTT ..... 3081  
182 ..... ThrIleLeuProValGlyTyrSerLeuGlu 191  
3080 ATAGCATTCTGTGCGCACAGCATCAAGCCCTTTTCATCT...TCAGT 3034  
192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208  
3033 TAAACCGCGCTCGTAGATAGTGAACCTGTGCGAAAGAACGACGTAATGC 2984  
208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225  
2983 CAGAAAGCCCTTCGTGCGAGCCGCCAGGCTACCAAAACCCCTTAGAAACA 2934  
225 InThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241  
2933 CTCGAA ..... ACCACGTCGCTTT 2914  
242 LeuSerGluGluTriThrValHisSerGlnGluAlaThrThrSerAlaTr 258  
2913 CGAACCGCCCAAGAACCTGCCCTC ..... 2889  
258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275  
2888 ..... GTACACTGG ..... 2880  
275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheYr 291  
2879 ...ACGTTAGAAACGTTCCCATTTATGACGCATATACAGAACTTTCAGT 2833  
292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu... 307  
2832 AATTGCGAGGGTTTCCATCTTCTGCTCTTCATTGCTTCACCTCAGCTTAA 2783  
308 ..... SerThrGluLysLeuAsnAsnSerThrGlyL 318  
2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCCG ..... AAGCGACTA 2739  
318 euGlnSerSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334  
2738 TTCCTTCAGTTTCCACGAAGTC ..... TCCTTCAAAATATCAAGGC 2695  
335 PheThrAspGlyProArgThrLeuArgSerLeuThrValSerLeuG1 351  
2694 ACCCTTGACCTCAACCTGCTCTGAAGCTCCAGAGTCCAAATGGTAACAA 2645  
351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368  
2644 CAACAGACCGAGTACTTTACGGATGAACGCTCTGGGC ..... 2607  
368 hrSerSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384  
2606 ..... ATAGACATAAGCGT 2593  
385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401  
2592 CTCCAAAACGTTGAAGTTATCTGTCACCATTCCTTACGATAGACGGAGA 2543  
401 rGluAsnGluPheGlyLeu ..... ThrSerL 410

2542 CTGCGTTTGAAGAAACAATCGCTACGTTCTGAAAGAAGAGCTTGAAGACC 2493  
410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426  
2492 CCGGGAGACTCCCTCACTGCTTTTATAAGTTATAGGAGCGAATACT 2443  
427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu ..... 440  
2442 CTTCCGCCCTATCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393  
441 ..... ThrValSerArgSerValAlaPro. 448  
2392 ACATATCATCAGGAGAAATATCAACCTTGTGTGG ..... 2358  
449 ...MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463  
2357 ACAACGCGACTCCCTGTA ..... 2340  
464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480  
2339 ..... ATCACATGGGTCTTTTGGCAGCTGACC ..... 2313  
480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497  
2312 ..GCCCAACGGCATACTCAAAGTTCTATTCTCGAATGTCCCTATAGTG 2265  
497 lySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThr 513  
2264 TTGAGCGCTAGGAATAGATCTGTGCGAACAGTCCGAAATCGAGGACTAT 2215  
514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529  
2214 TGTCTTAAAGGACACACTAGAACCA .....G 2189  
529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerA 546  
2188 CTACAACAGGTATGAACACACAGCATTTGCCACGAAACGACGGGCAAC 2139  
546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561  
2138 GACCTATACGTAACTCAGT .....TCCAGGTGACTTCTGAGCCCTAC 2096  
562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578  
2095 ACATAGTGAGATCATGGTAACACACACGCTATCCCTCATACTACTAGCTTA 2046  
578 rLeuThrAsnThrSerMetSerThrThr ..... 587  
2045 AACAGCTTCCCGACGAGTGGCTACATAGAAAGGTGAA .....AAA 2002  
588 ..... SerGlyGluAlaGlySer 593  
2001 CCGGCTTACGTT .....AGAGAAGGANTTTCATCATCTCATC 1964  
594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal... 609  
1963 CACACTGACCAACATTGTAGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914  
610 .AsnValThrAspAspMetGlyLeuVal .....SerA 620  
1913 GT .....ATCGAGTTCTTC 1900  
620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636  
1899 CAAGTAAACTTCACGTTTACCGTTACCAACACTCCACGGGAACAT ..... 1855  
637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653  
1854 ..... AAGAGCTATGTCGAAACACGACGCGCACCTTCCTTA 1821  
653 pH1sThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669

```
1820 TCATAGCCTCTGCTACCTGCTGGGGGATATA.....TCGTACACATC 1777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
670 SerIleThrAspAsnSerSerSerAspIleValGluSerSerThrSe 686
1776 CACATAAAACCCGCATCGCTAGAAATCGAGTTCTGGGATTCTCGAC 1727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 rTyrIleLysIleSerAsnSerHisSerGluTyr..... 698
1726 TACAGTAAGACAGACATTGAAGCCCTTCAGTTAGAAGTCA...CTG 1680
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 .....SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711
1679 GACAGTACGAA.....TCCTCCCCAGCTGTGGAAGAAAGAGAT 1639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu..... 724
1638 CTCATTGATTCCTCGCTTCACATCTTTAAGCTCTAGGACTGGATTGCATA 1589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr..... 737
1588 CATGACAGTTTCATGCGCCAGCTTTGACATGATACGTGAAGTCCGCCCA 1539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
738 .....ProThr 739
1538 ATATCGAAGAAGCACACATTACCACACCTCTCTGGAAGATTCCATTGAA 1489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 IleAsnMetProAsnThr..... 745
1488 ACAGCTCCGCATGGCATTAGCGAATGGTGACTACTCTCGACGAGTCGG 1439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
746 .....ServValValLeuAspThrAspAlaG 754
1438 AGAAGCTATCGAAGCTCTGGAAGAGTTCCCTCAGCTTCT..... 1397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
754 IuPheValSerAspSerSerSerSerSerSerSerSerSerSe 770
1396 .....CCCTTCCCGCGCTTCAGAAACACATCGCA 1366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis.Hi 786
1365 CCTATGTAGGGTTTTTCGGGAATCATTTGCTGCCCTTTGAGAAACATAT 1316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
786 sLeu.....PheSerIleLeuProSerThrArgAlaSerVal. 799
1315 CGACAAATTTGCTGTCAAATGGAAGCAATGCCGATTGCTCCCTTAAT 1266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSer 815
1265 ACCAGTCTATTGCTGATCTCGAAACGACGATCAATAGACCTCTCTTG 1216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816 ProSerProLeuProValSerLeuThrThrSer.....ThrSerAlaPr 830
1215 ATTAAGGCGATTGCGACGAGGCGCAAGATCTCCCTCCCAACGCTGCC 1166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847
1165 TATCGCCAGGAGGATAGCGGTTTTTGGGACATGGGGAGATAGACTCT 1116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863
1115 CCGCCCTTCCACCAAGGTACGCTCTTCATTGATGATGCGATGCTAG 1066
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880
1065 TATGTTGCTCGCGCAAACTAATAGCTTTAAGCATCTAAAG...CCAA 1019
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897
1018 GTTGATCAAGAAACCTATCGAAAG..... 993
897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913
992 .....GAAGATGAGAAGCAACGAGCTCTGTAAAT 964
```

```
914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928
963 TTCTCCCTCAGCCTCTCATTTGTTAAACCAACACACCCCTTTGGTGGT 914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
929 ....ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThr 944
913 ATACAAAGCCCTTTCCGGCATACTCCACAACCTCAATCGCGCGCAAGCA 864
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
944 hrThr.....IleLeuLysThrSerGlnProLeuMetThr 955
863 ACATCACACCCCTGGGTGATTTCACGATGAAGTCCCTCGATC..... 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
956 Thr.....ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970
821 .GTAGCCCTACCCCGCACAAATAGGAAGAATTTAACAGCTCTATCTTTGA 773
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
970 oIleAlaValGlnThrThrAlaGlyLysGln.....LeuSerLeuT 984
772 CGTAAATAATACCTCTCTCTAGGAGAGGCTTCATCCAG.....ATTTTA 729
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyLysSer 1000
728 GCCCAACGACGCTGTTTACATCAATAGGAACATTAAAGGCCCTGTT 679
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017
678 AAACAATTTCCCT.....TTCCGCAAGATGACAGCGTCTCTCA..... 639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034
638 .....TCACCCCCACCTTTTACTGAAGAAAGCCCAACGTCACG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1034 hrAlaGluTyrSerPro.....AlaSerArg.SerLeuGlyTh 1046
599 GCGGCTTCTCTCCCAACGACTGTTGGGACT.....CATTCGCTG 557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063
556 GCTCAGTGTCTCTGTTTGTAGGAAGTGTGAGTGGACGCCCAACAGCT 507
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 rOlysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079
506 GTTGGCTTAACAGCCCAATGTGACGTTGCCGTTTGGCTTCTGCG 462
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 LeuSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094
seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep.US-08-944-496-40
seq_documentation_block:
; Sequence 40, Application US/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF INVENTIONS: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
```







```
1115 CCGCCCTTCCACCAAGGTAGCTCTTTCATTTAGTGTGATGATGCTGTAG 1066
      ||| : : : : : ||| : : : : : ||
864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880
      ||| : : : : : ||| : : : : : ||
1065 TATGTTCTGTCGCGCAACCTTAATGACCTTTAAGCATTAAG...CCAA 1019
      ||| : : : : : ||| : : : : : |||
880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897
      ||| : : : : : ||| : : : : : |||
1018 GTGCATCAAGAAACCTATCCAAAG..... 993
      : : : : : ||| : : : : :
897 erLeuValSerLeuProThrGluSerThrLysAlaValThrAsnSer 913
992 .....GAAGATGAGAAGCAACGACGCTCTCTAAAT 964
      ||| : : : : : ||| : : : : : |||
914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu.... 928
963 TTCGCCCTCAGCTCTCATTTGTTAAACCAACACACCTTTGGTGGTC 914
      ||| : : : : : ||| : : : : : |||
929 ...ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944
913 ATACAAACGCCCTTTCGGCGCATCTCCACAACCTCAATCGCGCGCAAGCA 864
      ||| : : : : : ||| : : : : :
944 hrThr.....IleLeuLysThrSerGlnProLeuMetThr 955
863 ACATCACACCCCTGGGGCTGTATTACGATGAAGTCTCGATC..... 822
      ||| : : : : : ||| : : : : : |||
956 Thr.....ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970
821 .GTAGCCCTACCGCGCACATAGGAAGAATTTACAGCTCTACTCTTGA 773
      : : : : : ||| : : : : : |||
970 oIleAlaValGlnThrThrAlaGlyLysGln.....LeuSerLeuT 984
772 CGTAAATAACCCCTCTCTACGAGAGGCTTCATCCAG.....ATTT 729
      ||| : : : : : ||| : : : : : |||
984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000
728 GCCCAAGCAAGCTGTTCTTACATCAATAGGAACATTAAAGGCCCTGTT 679
      : : : : : ||| : : : : : |||
1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017
678 AAACAAATTCCT.....TTCGCGAAGATGACAGCGTCGCTCA.... 639
      : : : : : ||| : : : : : |||
1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034
638 .....TCACCCCCACCTTTTACTGAAGAAAGGCCAAACGTCAG 600
      ||| : : : : : ||| : : : : :
1034 hrAlaGluTySerPro.....AlaSerArg.SerLeuGlyTh 1046
599 GCGGCTTCTCTCCACGACTGTGTGGGACT.....GATTGCGCTG 557
      ||| : : : : : ||| : : : : : |||
1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlap 1063
556 GCTCCAGTGTCTCTGTTTTTGTAGAGTGTAGTGTGACGCCACACGCT 507
      : : : : : ||| : : : : : |||
1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079
506 GTTGGCTTAACAGCCCAATGTGAGCTTGGCTTGGGCTCTGCTG 462
      : : : : : ||| : : : : : |||
1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094
seq_name: /cgn2_6/ptodata/2/iaa/5b_comb.pep:US-08-682-517-15
```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-682-517-15

alignment_scores:
    Quality: 150.50          Length: 927
    Ratio: 0.358             Gaps: 45
    Percent Similarity: 45.307 Percent Identity: 19.957

alignment_block:
US-09-301-906-3 x US-08-682-517-15 ..

Align seg 1/1 to: US-08-682-517-15 from: 1 to: 1222

3784 GATGAGACGCTTAGTATTACCGGACTTTGACTATGACGAGGTGCTGG 3833
      ||| : : : : : ||| : : : : : |||
164 AspPheAlaLeuValPheLysArgThrValAspLysValGluGlyGluTh 180
3834 TCCATCTCGGACTCAGCATGAAGCGGTTCCCGGTGACGATACGATGAT 3883
      ||| : : : : : ||| : : : : : |||
180 rPro.....GluGluAlaAlaPheValLysAlaIleAsnAsnT 193
3884 CCACCTTCTAGTCTCAAGCTATGATGTTGCACAAATGTGCGCACGCTG 3933
      : : : : : ||| : : : : : |||
193 hrThrValGluValThrPheGluGluGluValThrAsnValGlnAlaLeu 209
3934 GGGATTAGCACCAACGG...GAAGTTACTGTGTGAAGAGAGACCCATT 3980
      : : : : : ||| : : : : : |||
210 AsnPheLysIleGluGlyLeuGluIleLysAsnAlaSerValLysGlnTh 226
3981 ACCTCGAAGCGTGCATACACTATGTCGAGAGAGAGGTGCGCGCTGTG 4030
      : : : : : ||| : : : : : |||
226 rAsnLysLysValValValLeuThrThrGluAlaGlnThrAlaAspLysG 243
4031 CAGCTGTGCGCGAAAGACAAGTGATCGTCGGT.....TCTGGTACC 4074
      ||| : : : : : ||| : : : : : |||
243 luTyThrValLeuThrLeuAspGlyGluThrIleGlyGlyPheLysGlyVal 259
4075 GCTGACGCTATG...GCTTTGTTGAAAGTGTGAAAAAGGTGTCGACGA 4121
      ||| : : : : : ||| : : : : : |||
260 AlaAlaValValProThrLysValGluLeuValSerSerAlaVal..... 274
4122 TGTCTTTCACCAACAGCTCTAGTGGGAAACGCTCGTGTGAGTTGAGTGG 4171
      : : : : : ||| : : : : : |||
275 .....GlnGlyLysLeuGlyGlnGluValLysValG 285
4172 ACGCAAA..... 4179
      : : : : : ||| : : : : : |||
285 lnAlaLysValThrValAlaGluGlyGlnSerLysAlaGlyIleProVal 301
4180 .....GGTTGCTCCCAAGAAAGCT 4199
      ||| : : : : : ||| : : : : : |||
302 ThrPheThrValProGlyAsnAsnAsnAspGlyValValPro...ThrLe 317
4200 GTCGGTGTAGGCGCGACACAAAGAGGGA..... 4230
      : : : : : ||| : : : : : |||
317 uThrGlyGluAlaLeuThrAsnGluGluGlyIleAlaThrTySerTyT 334
4231 ..AGAGCTGCAGATGTTAACACAGCACAAACGCGGTCAACGAAAGCGAC 4278
      ||| : : : : : ||| : : : : : |||
334 hrArgTyThrLysGluGlyThrAspGluValThrAlaTyThrAlaThrGlyAsp 350
```

4279 AGG..... 4281  
4351 ArgSerLysPheSerLeuGlyTyrValPheTrpGlyValAspThrIleLe 367  
4282 .....GAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTG 4321  
4367 uSerValGluGluValThrThrGlyAlaSerValAsnAsnGlyAlaAsnL 384  
4322 ATATTCGAAGGTCAACCCAGTCGAGGTACATGCTCAGAAA.....GAAGTG 4368  
4384 ystThrTyrLysValThrTyrLysAsnProLysThrGlyLysProGluAla 400  
4369 AAACAAGAAGTACCATGCGAGCTGTTTCGGCGCCACGCCAATCGTGA 4418  
4401 AsnLysThrPheAsnValGlyPheValGluAsnMetAsn....ValThrSe 416  
4419 TGAGAAACCCGCCCAAGTGTACGACTCGTGGTGTGAAGATAATGACA 4468  
4416 rAspLysValAla....AsnAlaThrValAsnGlyValLysAlaLeuGlnL 432  
4469 AGGGCAAGCCGTCGCTCATGTGGCTGAGAAAACACAGGTACAGTCGAG 4518  
4432 euSerAsnGlyThrAla.....LeuAspAlaAlaGlnIleThrAsp 446  
4519 CAGCCCAACAGAGGAGTTTGAGATCAATGAAGCAAGCCGCGTAACA 4568  
4447 SerLysGlyGluAlaThrPheThrValSerGlyThrAsnAlaAlaValTh 463  
4569 GCTTTGCATGTTTAGAACGCTGTTCTCGCGGTGTCAGCTGGATGTACA 4618  
4463 rProValValTyr.....AspLeuHisSerThrA 473  
4619 AGGACCGACTATCGCCACAGGCTTCTCAACGCATTTACCTTTGTCGAT 4668  
473 snAsnSerThrSerAsnLysLysTyrSerAlaSerAlaLeuGlnThr 489  
4669 AACTTGAAGGAGGAGTCGCGTCTTTCTCAAAAGCTGGTGGTGGTGA 4718  
490 .....AlaSerLysValThrPheAlaAlaLeuGlnAlaGluTy 502  
4719 TACCTAT.....AATGGTGTAGCCATGTTTTCATCAG 4750  
502 rThrIleGluLeuThrArgAlaAspAsnAlaGlyGlu..... 514  
4751 GTGGCGCTCGTCCCTAGAGGATATCTTAACGGCAATTAAAGTACCCAGC 4800  
515 .....ValAlaAlaIleGlyAlaThrAsn 522  
4801 GTCTTCGACCACTGTTTACTGCAGAGTACAAGATGGTGA..... 4842  
523 GlyArgGluTyrLysValIleValLysAspLysAlaGlyAsnLeuAlaTy 539  
4843 .....GGCGTACCATTCCACGCTGATGACGAGGAGTGC.... 4875  
539 sAsnGluIleValAsnValAlaPheAsnGluAspLysAspArgValIleS 556  
4876 .....TATCCATCAGATAACCTATCTTGACGGTC 4905  
556 erThrValThrAsnAlaLysPheValAspThrAspProAspThrAlaVal 572  
4906 AATCTCGTGGG...AAGGCA...AACTTCTGACTAAGTCAGG...AA 4946  
573 TyrPheThrGlyAspLysAlaLysGlnIleSerValLysThrAsnAspLy 589  
4947 GGGTGGTAAGGTACGCTCATAAAGCTAGCTTCGGGTGACTATTTCTTA 4996  
589 scLyGluAlaThrPheValIleGlySerAspThrValAsnAspTyrAlaT 606  
4997 TGCCTTGGCGTTTCAAAGGACGACTTGCATTCAGTAACTCCATCGAC 5046  
606 hrProIleAlaTrp.....IleAspIleAsnThrSerAsp 617  
5047 GAAGGCGGCATCAGTTTACGTTTACGTTACGGCAACTCGCGCGCTTTGGTGT 5096

618 AlaLysGln..... 620  
5097 AGCAGGATGTTCCAGTTAGCCGCGCGCTGTGCGATGAGAAGTCACCAG 5146  
621 .....GlyAspLeuAspGluGlyGluProL 629  
5147 GTGTTTCCAAC.....CAGCAACACAGACCCAGGTGCT 5181  
629 ysaValAlaAlaProIleSerTyrPheGlnAlaProTyrLeuAspGlySer 645  
5182 .....ACCAGAACATCACACAAA 5201  
646 AlaIleLysAlaTyrLysLysSerAspLeuAsnLysAlaValThrLysPh 662  
5202 ATCGGGGGCAAGCT.....CTATCTGAGGGAAGTG 5233  
662 eAspGlySerGluThrAlaValPheAlaAlaGluLeuValAsnGlnSerG 679  
5234 CTAGGGAAGTCAAGGG.....AGTCGACATACTCGATATGG 5271  
679 lylLysLysValThrGlyThrSerIleLysLysAlaThrTyrThrIle... 694  
5272 TCGCAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGCTCATATCC 5321  
695 .....TyrAsnThrGlyAlaAsnAspIleLysValAspAsnG 707  
5322 AGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAAGTGTT. 5370  
707 nValIleSerProAsnArgSerTyrThr...ValThrTyrGluAlaThrL 723  
5371 .....AAAACCGGACCTCTGAAGATGCCGCTCGTGGAGTACTTGAAGTAT 5415  
723 euSerSerThrGlyThr..... 728  
5416 CTGGCTATAGGCATTGAGAGGACATACAGGCGCTTCTTATGCTAGAAA 5465  
729 .....ValIleThrProAlaLysAs 735  
5466 TATTGCGCTCACTACCGCGAAGGT.....GTTCTGAAAGTACCTA 5506  
735 nLeuGluValThrSerValAspGlyLysThrThrAlaValLysVal.... 750  
5507 ATCAAGTTTATCAATCACTACCGCGCTTTCACGTTTACAGTCGGGCACA 5556  
751 .....IleAlaThrGlyIleAlaValAsnThrAspGlyLys 762  
5557 GATCTCATTTT.....CATTCACACACAGCGCTTGCCTGTGAG 5597  
763 AspTyrAlaPheThrAlaLysGluAlaThrAlaThrPheThrAlaThrAs 779  
5598 AGACCTACCGTACGTACTCATAGCTGAAAAGGTATCTTTACCACAGGCA 5647  
779 nGluValProAsn.....SerTyrThrGlyValAlaThrGlnPheA 793  
5648 AGATGTCACCGGTTAGTCTTGGCGCACAACTCTGTTCTGTCATCCGAC 5697  
793 snThrAlaAspSer.....GlySerAsn..... 800  
5698 GATATACTGTTTCCACCATGCCATTAATTTGATAGTGCCTGAAAGT 5747  
801 .....SerAsnSerIleTyrPheAlaGlyLysAsnProVa 812  
5748 CGCTCGATCGGCATGGTGGCAATCGTTTAAAGTCCTTC..... 5787  
812 lLysTyrAlaGlyValSerGlyLysThrTyrLysTyrPheGlyAlaAsnG 829  
5788 .....GAATATAAGTCTATAATGCT 5808  
829 lYasnGluValPheGlyGluAlaAlaTrpGluAlaLeuLeuThrGlnTyr 845  
5809 CCCCAGGTGCGGTAAAGCAGC.....AC 5834  
||| |||:|:|:|



4619 ACGAAGCGACTATCCCCACCAGGTTCCTCAACAGCATTTACCTTGTGCGAT 4668  
|||:::||| :::::||| :::  
503 sNAsnSerThrSerAsnLysLysTyrrSerAlaSerAlaLeuGlnThr 519  
4669 AACTTGAAAGGAGGAGCGCGCTCTTTTCAAAGCTGGGTGAGGGGTA 4718  
||| ||| |||:::||| |||  
520 .....AlaSerLysValThrPheAlaAlaLeuGlnAlaGluTy 532  
4719 TACCTAT .....AATGGTGGTAGCCATGTTTCATCAG 4750  
|||| |:::|||:::|||  
532 rThrIleGluLeuThrArgAlaAspAsnAlaGlyGlu..... 544  
4751 GGTGGCCCTCGGCCCTAGAGGATATCTTAACGGCAAATTAGTACCCAAGC 4800  
|||:::|||:::|||:::  
545 .....ValAlaAlaIleGlyAlaThrAsn 552  
4801 GTCTTCGACCACTGTTTAGTCAGAAGTACAAGATGGTGGGA..... 4842  
|||:::|||:::||| ||| |||:::|||  
553 GlyArgGluTyrlsValIleValLysAspLysAlaGlyAsnLeuAlaLy 569  
4843 .....GGCGTACCATTCCACGCCTGATGACGAGGAGTGC.... 4875  
||| ||| |||:::||| |||  
569 sAsnGluIleValAsnValAlaPheAsnGluAspLysAspArgValIle 586  
4876 .....TATCCATCAGATAACCTCATCTTCACGSGTC 4905  
586 eThrValThrAsnAlaLysPheValAspThrAspProAspThrAlaVal 602  
4906 AATCTCGTGGG...AAGGCA...AACTTCTCGACTAAGTGCAGG...AA 4946  
603 TyrPheThrGlyAspLysAlaLysGlnIleSerValLysThrAsnAspLy 619  
4947 GGTGGTAAAGTCATGTCATAAACGTAGTTCGGGTGACTATTTCTTA 4996  
||| ||| |||:::|||:::|||:::  
619 sgLyuAlaThrPheValIleGlySerAspThrValAsnAspTyrrAlat 636  
4997 TGCCCTTCGGTTTTCAAGAGGCGCAC TTGCATTTCAGTAAACTCCATCGAC 5046  
||| |||:::|||:::|||:::|||  
636 hrProIleAlatr.....IleAspIleAsnThrSerAsp 647  
5047 GAAGGGCGCATCAGTTTGACGTTCCAGGCGCAACTCGCGCGCTTGTGTGT 5096  
:::  
648 AlaLysGln..... 650  
5097 AGCGAGGATGTGCAGTTAGCGGGGGGTGTCGGATGAGAAGTACCAG 5146  
||| |||:::|||:::|||  
651 .....GlyAspLeuAspGluGluProL 659  
5147 GTGTCTCAAAC.....CAGCAACACAGACGCCAAGTGC 5181  
::: ||| |||:::|||:::|||  
659 ysAlaValAlaProIleSerTyrrPheGlnAlaProTyrrLeuAspGlySer 675  
5182 .....ACCAGAACATCACACAAAA 5201  
676 AlaIleLysAlaTyrrLysLysSerAspLeuAsnLysAlaValThrLysph 692  
5202 ATCGGGGGCGAAGGCT.....CTATCTCGGGAGTGC 5233  
692 easpGlySerGluThrAlaValPheAlaGluLeuValAsnGlnSerg 709  
5234 GTAGGGAAGTCAAGGG.....AGTGCACATACTCGATATGG 5271  
|||:::||| |||:::|||:::|||:::|||  
709 lyLysLysValThrGlyThrSerIleLysLysAlaThrTyrrThrIle... 724  
5272 TC CGAACAAAGATATTCGTTAGGAAGTGTGAGTGGCTCAGGCGCTGATATCC 5321  
||| |||:::|||:::|||:::|||  
725 .....TyrAsnThrGlyAlaAsnAspIleLysValAspAsnGl 737  
5322 AGTGATGCTCTTGAACCTGACTACCCCCAAYGACATTTGAAGTGGTT.. 5370  
|||:::|||:::|||:::|||:::|||  
737 nValIleSerProAsnArgSerrTyrr.....ValThrTyrrGluAlaThrL 753

5371 .....AAACCGGGACCTCTCTGAAGATGCCGCTGGGAGTACTTGAAGTAT 5415  
 753 euSerSerThrGlyThr .....ValIleThrProAlaLysAs 765  
 5416 CTGGCTATAGGCATTGAGAGGACATACAGGCGCTTGCTTATGGCTAGAAA 5465  
 759 .....ValIleThrProAlaLysAs 765  
 5466 TATTGCCGTCTACCTACCGCGAAGGT .....GTTCTGAAGACTACCTA 5506  
 765 nLeuGluValThrSerValAspGlyLysThrAlaValLysVal ..... 780  
 5507 ATCAAGTTTATGAATCACTACCGGCTTTACAGTTTCAAGTCGGGCACA 5556  
 781 .....IleAlaThrGlyIleAlaValAsnThrAspGlyLys 792  
 5557 GATCTCATTTTT .....CATTCACACAAAGACGGCTTGCTGTGCAG 5597  
 793 AspTyrAlaPheThrAlaLysGluAlaThrAlaThrPheThrAlaThrAs 8099  
 5598 AGACCTACCGTACGTACTCATGAGTGAAGAAAGTACTCTTACCAAGGCGCA 5647  
 809 nGluValProAsn .....SerTyrThrGlyValAlaThrGlnPheA 823  
 5648 AGATGTGCAACGGGTGGTAGCTTTGGCGCAATCTGTCGTATGCGAC 5697  
 823 snThrAlaAspSer .....GlySerAsn ..... 830  
 5698 GATATACTGGTTTCCACGATGCCATTAAATTGATAGGTGCACCTGAAGT 5747  
 831 .....SerAsnSerIleTyrPheAlaGlyLysAsnProVa 842  
 5748 CGCTCGATCGGCATGGTGGCGAATCGTTTAAAGTCCTTC ..... 5787  
 842 LysTyrAlaGlyValSerGlyLysThrTyrLysTyrPheGlyAlaAsnG 859  
 5788 .....GAATATAAGTCGTATAATGCT 5808  
 859 LysAsnGluValPheGlyGluAlaAlaTyrGluAlaLeuLeuThrGlnTyr 875  
 5809 CCCCAGTGGCGGTAAACAGCAGC .....AC 5834  
 876 AlaThrGluGlyGlnLysValThrIleSerTyrAsnValAspGlyAspThr 892  
 5835 GTTAGTGGCAATTCGTTAAAGTCACCCAAATAGCACAGCC ..... 5874  
 892 rValThrPheLysValIleSerAlaValAsnSerSerThrGluAlaIle 909  
 5875 .....ACC 5877  
 909 ysProValAlaProThrThrProAlaAlaProThrThrGlyAlaLeuThr 925  
 5878 ATTACGGCTAATCGGGAAGTCTGAGGACATAAATATGCGGTGAAGAA 5927  
 926 LeuThrProAlaAlaGlyGlyLeuValAspLeuThrThrAlaThrAsnThr 942  
 5928 GAGATCCGAATTG .....GAAGGTCTCAAC .....AGTGCTACCACAG 5968  
 942 rLeuGlyIleSerLeuAlaAspAlaAspLeuAsnValSerAlaThrThrV 959  
 5969 TTAACCTCCAGGCTGTAACCTTTATTCGTCAGGGAATGCTATAAAGGGTT 6018  
 959 aAspThrAlaThrValSer ..... 965  
 6019 TTGCTGGATAGGTGCACATGATGCATCAAGGCTTACTACAACTAGCGT 6068  
 966 LeuLysAspSerAlaAsn .....AsnSerLeuSerLeuThrLe 978  
 6069 CTTCGCACCGCGCTCGGAAGCGCTCTTT 6099  
 978 uValGluThrGlyAlaAsnThrGlyValPhe 988

seq\_name: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep:US-08-362-525-22





```
2903 AGAAGACCTGCCCTCGTACAGTGGAGCTTAGAAACGTTCCCAATTATGAA 2854
      ::::::::::: ||| :::::::::::
806 ...ThrLysThrValValThrSerSerLeuSerArgSerAsnHisAlaG1 821
      ::::::::::: ||| :::::::::::
2853 CGCATATACAGAACTTCA.....GTAATTGGCAGGGGTTTTCATCTT 2810
      :::: ||| :::: ||| ||| ||| ::::
821 uThrGlnThrAlaSerAlaThrAspValIleGlyHis..... 833
2809 CTGTCCTTCATTGCCCTTCATGACCTTAATTGAATCTTCCACGATAGCCCTA 2760
      :::: ||| ||| :::: ||| ||| :::: ||| ||| ::::
834 .....SerSerSerValValSerVal 840
2759 GAGAAAGAGG.....CGCAAGCGCACTATTCTTCAGTTTCCCTACCGAAGGT 2716
      ::::::::::: ||| ||| ||| ||| ||| ::::
841 SerGluThrGlyAsnThrLysSerLeuThrSerSerGlyLeuSerThrMe 857
2715 CTCCTTCGAATATCAAGGACCCCTGAGCTCC 2682
      ||| :::: ||| ||| ||| |||
857 LserGlnGlnProArgSerThrProAlaSerSer 868
```



OM of: US-09-301-906-3 to: PIR\_66:\* out\_format : pfs  
 Date: Jan 31, 2001 3:49 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -DEV=xlpl  
 -MODEL=framed\_n2p.model -DB=PIR\_66 -Q/cgnt2\_1/USPTO\_spool/US09301906/runat\_31012001\_153035\_9612/app\_query.fasta\_1.6881  
 -DB=PIR\_66 -Q/cgnt2\_1/USPTO\_spool/US09301906/runat\_31012001\_153035\_9612/app\_query.fasta\_1.6881  
 -MINMATCH=0.100 -LOOPCL=0.100 -LOOPT=0.000 -GAPOP=4.000  
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
 -DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi  
 -LIST=45 -DOCALIGN=200 -OUTSCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODE=LOCAL -OUTPCT=pfs -NORM=ext -MINLEN=0  
 -MAXLEN=200000000 -USER=US09301906 -CGN1\_1\_784 -NCPU=6 -ICPU=3  
 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
 Query: US-09-301-906-3  
 Query length: 6714  
 Database: PIR\_66\*  
 Database sequences: 195891  
 Database length: 67900655  
 Search time (sec): 527.440000

score_list:	Strd	Orig	ZScore	Escore	Len	Documentation
pir2:JA4282	+	629.00	718.37	2.6e-33	1729	! fusion protein la/lb - sugar b
pir1:JQ1734	+	414.00	462.66	4.6e-19	1718	! genome polyprotein - shallot v
pir2:JQ0096	+	404.00	451.97	2.0e-18	1547	! hypothetical 176K protein - pa
pir2:JQ0096	+	394.00	438.94	9.6e-18	1707	! hypothetical protein 1 - grape
pir2:JQ0096	+	391.50	435.95	1.4e-17	1707	! hypothetical protein 1 - grape
pir2:S01865	+	338.50	372.91	4.6e-14	1707	! genome polyprotein - tobacco
pir1:WMFM12	+	329.50	366.98	1.5e-13	1126	! 125K protein - alfalfa mosaic
pir2:T45025	+	318.00	340.04	1.5e-12	3570	! mucin MUC5B, tracheobronchial
pir1:WMFM8T	+	309.00	338.45	4.0e-12	1615	! 180K protein - tomato mosaic
pir2:S03701	+	306.00	333.85	6.7e-12	1766	! 141K protein - pea early brown
pir1:WMFM2	+	297.00	324.20	2.5e-11	1611	! 183K protein - pepper mild mot
pir2:JQ1896	+	294.50	320.98	3.7e-11	1646	! 186K protein - cucumber green
pir1:PIBVB	+	291.00	316.49	6.4e-11	1694	! hypothetical 191.1K protein -
pir2:JQ2183	+	281.00	303.36	3.1e-10	1885	! RNA la protein - broad bean mot
pir1:PIVXPJ	+	277.00	305.97	4.2e-10	993	! RNA 1 protein - cucumber mosaic
pir1:PIBVA	+	275.50	304.05	5.3e-10	1005	! RNA 1 protein - peanut stunt v
pir1:PAVBBS	+	273.50	300.23	7.7e-10	1139	! alpha-a protein - barley stripe
pir1:PIVXCM	+	273.00	301.21	7.8e-10	993	! RNA 1 protein - cucumber mosaic
pir1:WMFMGM	+	272.50	295.08	1.1e-09	1608	! 183K protein - tobacco mild g
pir2:S25345	+	270.50	292.69	1.4e-09	1609	! probable membrane protein YCR0
pir2:E71392	+	269.00	296.45	1.4e-09	993	! ATP-dependent RNA helicase - cu
pir1:WMFM18	+	267.00	288.48	2.4e-09	1615	! 183.3K protein - tobacco mosai
pir1:JQ2144	+	264.50	285.50	3.6e-09	1616	! 183K protein - tomato mosaic v
pir2:S26358	+	264.00	289.16	3.2e-09	1116	! hypothetical protein, 126K - t
pir1:PIBVCV	+	263.00	289.34	3.6e-09	991	! RNA 1 protein - cucumber mosaic
pir1:JA5353	+	261.00	279.57	6.6e-09	1884	! genome polyprotein - apple chl
pir2:T34434	+	261.00	277.63	7.1e-09	2232	! hypothetical protein K06A9.1a
pir2:JQ2169	+	253.50	278.01	1.5e-08	993	! RNA 1 protein - cucumber mosai
pir1:PIBVCC	+	252.00	276.64	1.9e-08	958	! la protein - cowpea chlorotic m
pir2:S49432	+	252.00	274.92	2.0e-08	1112	! replicase 126K - odontoglossum
pir2:S57180	+	241.50	261.94	1.0e-07	1161	! probable membrane protein YJRI
pir1:S48478	+	240.00	258.28	1.4e-07	1367	! glucan 1,4-alpha-glucosidase
pir2:T30826	+	239.50	252.28	1.9e-07	2187	! nascent polypeptide-associated
pir1:PIVXTA	+	237.00	258.38	1.9e-07	993	! RNA 1 protein - tomato aspermy
pir2:S48699	+	236.00	251.70	2.8e-07	1601	! 178K protein - tobacco mosaic
pir2:S59310	+	224.50	242.29	1.3e-06	1104	! probable membrane protein YMR3
pir2:T34513	+	219.00	222.47	5.4e-06	3507	! hypothetical protein 27883.1 -
pir2:S65053	+	218.00	230.32	4.3e-06	1597	! genome polyprotein - Chikungia
pir2:AS5577	+	209.50	219.97	1.6e-05	1630	! ascites stialogycoprotein 1 -

pir2:T34369	-	206.00	214.81	2.8e-05	1777	! hypothetical protein T19D12
pir2:T42233	+	203.00	212.53	4.2e-05	1589	! submaxillary mucin 1 - bovi
pir2:A41258	-	198.50	216.19	5.8e-05	725	! a-agglutinin core protein AG
pir2:S25370	-	195.50	205.86	0.0001	1306	! MSB2 protein - yeast (Sacch
seq_name: pir2:A49282						
seq_documentation_block:						
fusion protein la/lb - sugar beet yellows virus (isolate BYV-U) (fragments)						
C:Species: sugar beet yellows virus, SBV						
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 17-Mar-1999						
C:Accession: A49282						
R:Agranovsky, A.A.; Koonin, E.V.; Boyko, V.P.; Maiss, E.; Froetschl, R.; Lunina, N.A.						
Virolgy 196, 311-324, 1994						
A:Title: Beet yellows closterovirus: complete genome structure and identification of						
A:Reference number: A49282; MUID:94082464						
A:Accession: A49282						
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual						
A:Molecule type: mRNA						
A:Residues: 1-1729 <AGR>						
A:Cross-references: GB:X73476						
A:Note: this protein is predicted to result from +1 ribosomal frameshifting during tr						
alignment_scores:						
Quality: 629.00						Length: 1862
Ratio: 0.834						Gaps: 76
Percent Similarity: 40.494						Percent Identity: 20.838
alignment_block:						
US-09-301-906-3 x A49282						
Align seg 1/1 to: A49282 from: 1 to: 1729						
1285 TTCTCTCCATTTGGACACGAAATTTGGCATATGTTTCTCAAGGGA 1334						
4 LeuAlaSerValGluLysAspLeuLeuAspPheLysAspGluLeu 20						
1335 CGCAATGATTCGGGAAACCCCTCACATAGTCGATGTGTTTCTGAAGC 1384						
20 sSerLeuSerLysGluLysArgSerValThrValProPheThrMetClyG 37						
1385 CGCGGAAAGGAGAGCTCAGGAGACTCTTCCAGAGCTTTCCAGACTTTCGATACAG 1434						
37 LuAlaValGlnSerGlyLeuThrArgAlaThrProGlnPheAsnLeuSer 53						
1435 TTCTCGCATTCGGTAGGAGTACCTACCCATTCGTAATGCCATCGCGAG 1484						
54 PheThrHisSerValThrSerAspHisProAlaAlaAlaGlySerArgLe 70						
1485 CTGTTTCAATGGAATCTTTTCCAGAGGAGTGTGTAATGTGTGTTCTTCG 1534						
70 uLeuGluAsnGluThrLeuAlaSerMetAlaLysSer...SerPheSerA 86						
1535 ATATTGGGGGAGCTTCACGTATCATGTCTCAAGCTGCCCATGTGGAAGTGT 1584						
86 sPheGlyGlyCysProLeuPheHisLeuLysArgGlySerThrAspTyr 102						
1585 CATGTATGCAATCCATCTAGACGTTAAAGATGTAAGCGGAGAGATCAA 1634						
103 HisValCysArgProIleThrAspMetLysAspAlaGlnArgArgValse 119						
1635 TGAGATCTCTTTTCTTCCACAGCTGGGGAGATTCCTACGTCTCGAGTG 1684						
119 rArg.....GluLeuGlnAlaArgGlyLeuValGluAsnLeuSerArg 134						
1685 ACCTTCTACTGAGCGGCTTCAAGTCTGTGTCTTACTGTAGTCGAGAA 1734						
134 LuGlnLeuValGluAlaGlnAlaArg...ValSerValCysProHisThr 149						
1735 TCGCAACTCGCATCTCTAGAGCGGATCGCGGTGTTTATGCTGATGTA 1784						
150 LeuGlyAsnValValLysSerAspValLeuLeuMetValGlnValty 166						



515 gLeuArgAsnPheValAlaTyrPhe.....ValGluA 526  
524 CACATGCTCTTCCACGGCATTAAATGTTTCTTTGGTTAGGACATTG 3573  
526 rgArgLeuAlaSerLysArgLeuLysThrPheIleGlyLeuAlaArgLeu 542  
5374 CTTGGGGGAGGGCTAGCGAATACTTGGAGCTTTGGAGGCTTCTCCAA 3623  
543 .....PheAspAsnPheSer.. 547  
3624 TAATGGTGCCTGTTCCGAGGTTGTTGGCAGGGAAGAGTGTGAGCT 3673  
548 .....LeuThrSerValValTyrLeuLeuGlnGluTyrAspS 560  
3674 CATTGTTACTGCTTACGCTAGGAGTATCTTTGTATCATAGGGGCTTG 3723  
560 erValLeuAsnAlaPhe...IleAspValGluLeuLeu 572  
3724 CTTAAGCAGCACCATACCTCAACTTGCTTACGTCCACCGGTAGAGGGAG 3773  
573 LeuAsn..... 574  
3774 GAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACG 3823  
574 ..... 574  
3824 AAGGTGCTGGTCCATCTGGGACTACGATGAAGCGGTTCCCGGTGACGAT 3873  
574 ..... 574  
3874 AACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCAAAATGT 3923  
575 .....SerGlySerValAsnValLeuProLeuValSerTrpVa 587  
3924 GCGCGAGCTGGGGATTAGCACCAACCGGGAAGTTACTGTGAAGAAGACA 3973  
587 larg.....GlySerLeuThrLysLeuAlaGlu. 596  
3974 CCCATTCACTCGAAGCGTGAATACACTTATGTCGAGGAGAGGTTGCC 4023  
596 ..... 596  
4024 CCCTCTGCAGCTGTGCGGAAGACAAAGTGATCCGTCGGGTTCTGGTAC 4073  
597 .....AlaIleVal.....GlySerGlyPh 603  
4074 CGCTACGCTATGCGTGTGTTGTAAGTGTGAAGGAGGTGTCGAGGATG 4123  
603 eala.....SerPheLeuGlyArgMetCysCysArgValSerAspt 617  
4124 TCTTTACCAACAGTCTAGTGG.....GAAACGGCTCGTGAGGTTGAG 4167  
617 rpCysSerSerSerAsnAlaGlyCysAsnPheMetSerProValArg 633  
4168 GTGGACGCAAGGGTGTCTCCAGAAACCGTCGTCGGTAGGCGCCGAC 4217  
634 ThrLysGlyLysPheValProProSerSerGlySerThrAlaSerMe 650  
4218 ACAAGAGGGGAAGCTGACAGTGAAGTGAACACAGCACAAACCGCGGTCA 4267  
650 tTyrGluArgLeuGluAlaLeuGluSerAspIle..... 661  
4268 ACGAAGGCGACAGGACGAGTACAGTCCAGTCTTGTGATGTTCCGCACAG 4317  
662 .....ArgGluHisValLeuSerThrCys..... 669  
4318 GCTGATATTCAAAGGTACCCAGTCCGAGGATGCTCAGAAAGAGT 4367  
670 .....ArgValGlySerAspGlu..... 675  
4368 GAAACAAGAGTACCATTGGCGACTGTTTCGGGCGCCACCGCAATCGTCG 4417

675 ..... 675  
4418 ATGAGAAACCCGCCCAAGTGTACGACTCGTGTGAAGATAATTGAC 4467  
675 ..... 675  
4468 AAGGGCAAGCGCTCGCTCATGTGGCTGAGAAAAACAGGTACAAGTCGA 4517  
676 .....GluGluG1 678  
4518 GCAGCCCAACACAGAGGAGTTTACGATCAATGAAGCAAGCGCGGTAAC 4567  
678 uArgProLysGlu..... 682  
4568 AGCTTTGCATGTTTAGAACGTGTTCTCGGGTGTGACGTGGATGTGTAC 4617  
682 ..... 682  
4618 AAGCAAGCGACTATCGCCACCGAGTTCTCAACGCGATTTACCTTTGTCGA 4667  
682 ..... 682  
4668 TAACTTGAAGGAGGAGTCCGGTCTTTTCTCAAAGCTGGGTGAGGGT 4717  
683 .....V 683  
4718 ATACCTATAATGGTGTAGCCATGTTTTCATCAGGGTGGCTCGTCCCTA 4767  
683 alThrGluProGlyIleGluHisThrSer..... 692  
4768 GAGGATATCTTACGGCAATTAAAGTACCCAGCGTCTTCGACCACCTGTTT 4817  
693 GluAspValVal.....ProIleArgSerHis..... 701  
4818 AGTGCAGAAGTACAAGATGGTGGAGCGGTACCATTCCACGCTGATGACG 4867  
702 .....SerGlnProLeuSerGlyGlyGlyCysSerTyrSerGluAspA 716  
4868 AGAGTGTCTATCCATCAGATAACCTATCTTTGACGGTCAATCTCGTGGGG 4917  
716 rgGlu.....GluAsnGlu..... 720  
4918 AAGGCAAACTTCTCGACTAAGTCCAGGAAGGGTGGTAAGGTCATGTCAT 4967  
721 ArgAlaAsn..... 723  
4968 AAACGTAGCTTCGGGTGACTATTTCTTATGCTTCCGCTTTTCAAAGGA 5017  
724 .....LeuLeuPro..... 726  
5018 CGCACTTGCAATTCAGTAAACTCCATCGACGAGGCGCATCAGTTTCACG 5067  
727 ..HisValSerLysIleValSerGluArgGly...LeuGluThrAla 741  
5068 TTCAGGGCAACTCGGCGCTCTTTGTGTAGGACGAGTGTTCAGATTAGC 5117  
742 ArgArgAsnLysArgThrLeuHisGlyValSerGluPheLeuAsnAlaI1 758  
5118 CGCGCGCGTTCGGATGAGAAGTCAACGAGTGTTCCAAACCAACCAACCAC 5167  
758 eAsnThrSerAsnGluGlnProArgProIleIleValAspHisSerProG 775  
5168 AGAGCCCAAGGTGCTACGAGAACAATCACACCAAAATCGGGGCAAGGCT 5217  
775 luSerArgAlaLeuThrAsnSerVal..... 783  
5218 CTATCTGAGGGAAGTGGTAGGAA.....GTCAAGGGAGGTCGAC 5258  
784 .....ArgGluPheTyrTyrLeuGlnGluLeuAlaLe 794  
5259 ATACTCGATATGTCGAA...CAACATTACGTTAGGAAGTGTGAGTGGC 5305  
794 upheGluLeuSerCysLysLeuArgGluTyrTyrAspGlnLeuLysValA 811

5306 TCAGGCTGATAATCAACAGTATGGCTCTTGAACCTGACTACACCCCAATG 5355  
::: ::::: ::::: ::::: |||  
811 laAsnPheAsnArgGlnGluCysLeuCysAspLysAsp..... 823  
5356 ACATTGGAAGTGGTTAAACCGGGACCTCTCAAGATGCCGTCGGGAGTA 5405  
||| ::::: ||||| |||  
824 .....GluAspMetPheVal..... 828  
5406 CTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGCGTGTGCTTA 5455  
828 ..... 828  
5456 TGGCTAGAAATATTGCCGTCACCTACCGCGGAAGGTGTTCTGAAAGTACCT 5505  
::: ::::: ||||| :::::  
829 .....LeuArgAlaGlyGlnGlyValValSerGlyArg 839  
5506 ANTCAAGTTTATGAATCACTACCGGGCTTTCAGGTTTACAAGTCGGGCAC 5555  
||| ::::: ||||| |||  
840 Asn.....SerArgLeuProLeuLysHisPhe.....LysGlyIle 851  
5556 AGATCTCATTTTTCATTCAACACAGACGGCTTGGCTGTGAGACACCTAC 5605  
::: ::::: ||||| |||||  
851 sGluPheCysPheArgSer.....GlyGlyLeu.....ValP 862  
5606 CGTACGTACTCATAGCTGAAAGAGGTATCTTTACCAAGGCAAGATGTC 5655  
||| ||||  
862 roTyr..... 863  
5656 GACGGCGGTAGCTTTCGGGCGACAACTGTTCTGATCGCAGCATATACT 5705  
||| ::::: ||||| |||  
864 .....AspGlyThrSerArgValAspThrIle.. 872  
5706 GGTTTTCACGATGCCATTAATTCATAGCTCCACTGAAAGTCGCTCCAT 5755  
||| ||||| ::::: |||||  
873 ....PheHisThrGlnThrAsnPheValSerAla...AsnAlaLeuLeus 887  
5756 GCGGATGGTGGCGGAATCGTTTAACTCCTTCGAATATAAG..... 5796  
||| ::::: ||||| ::::: |||  
887 erGlyTyrLeu.....SerTyrArgThrPheThrPheThrAsnLeuSer 901  
5797 .....TGCATTAATGCTCCCGCAGGTGGCGGTAGACAGCAC 5834  
||| ::::: ||||| |||||  
902 AlaAsnValLeuLeuTyrGluAlaProPro...GlyGlyLysThrThr 917  
5835 GTTACTGCACCAATTCGTTAAG.....TCACCCCAATAGCACGCCACCA 5878  
||| ::::: ||||| ::::: |||  
917 rLeuIleLysValPheCysGluThrPheSerLysValAsnSerLeuIleL 934  
5879 TTACGGCTAATGTGGGAAGTCTCTGAGGACATA.....AAT 5913  
::: ||||| ::::: ||||| |||||  
934 euThrAlaAsnLysSerSerArgGluGluIleLeuAlaLysValAsnArg 950  
5914 ATGGCGGTGAGACAGAGATCCGANTTTTGAAGTCTCACAGTGCTAC 5963  
||| ::::: ||||| |||||  
951 IleValLeuAspGluGlyAspThrProLeuGlnThrArgAspArgIleLe 967  
5964 CACACTTAACCTCCAGGCTGTTAACTTTATCGTCAGGGGAATGTATAAA 6013  
||| ::::: ||||| ::::: |||  
967 uThrIleAspSerTyrLeuMetAsn.....AsnArgGlyLeuThrCysL 982  
6014 GGGTTTTG...GTGATGAGGTGCACATGATGCATCAAGCTTTACTACAA 6060  
::: ||||| ::::: ||||| |||||  
982 ysValLeuTyrLeuAspGluCysPheMetValHisAlaGlyAlaVal 998  
6061 CTAGCGCTCTCGCAACCGCGCGTCGGAAGCCCTCTTTTGGAGACAT 6110  
::: ||| ::::: ||||| ::::: |||  
999 AlaCysIleGluPheThrLysCysAspSerAlaIleLeuPheGlyAspSe 1015  
6111 AAATCAGATACCATTCATAACACGGGACAGGTGTTTAGGATGGATTCGT 6160  
::: ||||| ::::: |||||  
1015 rarGlnIleArgTyr.....GlyArgCysSerGluLeuAspThrA 1029

6161 CTGTT.....TTTGTTCCAAAGAAAGCGCTTGTGA 6192  
||| ||||| ||||| |||  
1029 laValLeuSerAspLeuAsnArgPheVal...AspAspGluSerArgVal 1044  
6193 TACACTTCTAAATCGTACAGGTGCCGTTAGATGTTTCTACTTCTTGTGC 6242  
||| ::::: ||||| ||||| |||||  
1045 TyrGlyGluValSerTyrArgCysProTyrPaspValCysAlaTrpLeu 1061  
6243 CTCATGACCGCTAAGGGAACCGAAAGTGTTCACCTCGAAAGGTCGTT. 6291  
||| ::::: ||||| ::::: |||||  
1061 rThrPhe.....TyrProLysThrValAlaT 1070  
6292 .....AGCGTAAAGGACAAACACAGTAGTAGATCGCTG 6324  
::: ||||| ::::: |||||  
1070 hrThrAsnLeuValSerAlaGlyGlnSer.....SerMet 1081  
6325 TCCAAAGCCCAATTTGGAAACACACATGACGTAGCTGAAATAAACGCTGA 6374  
::: ||| ||| ::::: ||||| |||||  
1082 GlnValArgGluIleGluSerValAspAspValGluTyrSerSerGluPn 1098  
6375 CGTGACTTGTGCATGACCCAGTTGGAGAAGTCGGATATG..... 6414  
||| ||||| ||| ||| ||||| ::::: |||||  
1098 eValTyrLeuThrMetLeuGlnSerGluLysLysAspLeuLeuLysSerP 1115  
6415 .....AAGAGTCTGTTGAAGGAAAGAAAGAACACACAGTGATGACA 6459  
||| ||||| ::::: ||| ||| |||||  
1115 heGlyLysArgSerArgSerValGluLysProThr...ValLeuThr 1130  
6460 GTGCATGAAGCACACGGGAAACATTCAGTGTGTTGTTAGGAC 6509  
||| ||||| ||||| ::::: ||| ||| |||||  
1131 ValHisGluAlaGlnGlyGluThrTyrArgLysValAsnLeuValArgTh 1147  
6510 GAAGAAACCGCATGACTCCCTATTCTACATAAACACCGCATATCTGTTG 6559  
||| ||||| ::::: ||| ::::: ||||| |||||  
1147 rLysPheGlnGluAspPropPheArgSerGluAsnHisIleThrValA 1164  
6560 GTTGTCTGAGACACACACGCTCACTGTTTATGCGCTCTGAGCTCAAG 6609  
||| ||||| ||||| ::::: ||||| ||||| |||||  
1164 laLeuSerArgHisValGluSerLeuThrTyrSerValLeuSerSerLys 1180  
6610 TTGGACGATAAGGTCGGCACATATATTAGCGACGCG 6645  
||| ||||| ::::: ||| |||  
1181 ArgAspAspAlaIleAlaGlnAlaIleValLysAla 1192  
seq\_name: pir1:JQ1734  
seq\_documentation\_block:  
genome polyprotein - shallot virus X  
N;Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase  
N;Contains: RNA-directed RNA polymerase (EC 2.7.48)  
C;Species: shallot virus X  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C;Accession: JQ1734  
R;Kanyuka, K.V.; Vishnichenko, V.K.; Levay, K.E.; Kondrikov, D.Y.; Ryabov, E.V.; Zavr  
J. Gen. Virol. 73, 2553-2560, 1992  
A;Title: Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal cistron cio  
A;Reference number: JQ1734; MUID:93019008  
A;Accession: JQ1734  
A;Molecule type: genomic RNA  
A;Residues: 1-1718 <GB>  
A;Cross-references: GB:M97264; NID:g295078; PIDN:AAA47787.1; PID:g295079  
C;Keywords: ATP; nucleotidyltransferase; P-loop; RNA biosynthesis; RNA replication  
F;915-922/Region: nucleotide-binding motif A (P-loop)  
F;978-983/Region: nucleotide-binding motif B  
F;921/Binding site: ATP (Lys) #status predicted

alignment\_scores:  
Quality: 414.00 Length: 1025  
Ratio: 0.828 Gaps: 51  
Percent Similarity: 48.780 Percent identity: 22.244



```

... ||| |||||: |||:
885 l1euProLysLeuAspArgAspLeuLeu.....SerSerTrpValAlaL 900
5588 TG.....CGTCTGAGACACTACGTTACTCATAGCTCAAAA 5628
900 euAlaGluThrThrTrpArgGluValAlaValLeuAlaIleHisGlyAla 916
5629 GGTATCTTTACCAAGGGCAAGATGTCGACGCGGTGGTAGCTTTGGGGCA 5678
917 GlyGlyAlaGlyLysSerArgAlaLeuGlnGluLeuLeuArgSerSerPr 933
5679 CAATCTGTTCGTATCGGACGATATACGTGTTTCCACGATGCCATTAAAT 5728
933 oGluLeu.....AlaAspSerIleAsnIleValValProThrIleAsnL 948
5729 TCATAGGTGCACCTAAAGTCCCTCGATCCGCGCATGCGTGGCGAATCGTTT 5778
948 euAlaAsnAspTrpLysAlaLysLeuProGlnMetAspProArgArgVal 964
5779 AAGTCCTTCGAATATAAGTGTCTAATAATGCTCCCGCAGGTGGCGTAAGAC 5828
965 MetThrPheGlnLysAlaCysGluArgGlu.....CysLysSe 977
5829 GACGACGTTAGTGACGAATTCGTTAAGTACCCCAATAGCACACCCACA 5878
977 rValThrIlePheAspAspTyrGlyLysLeuPro.....AlaGlyPheV 992
5879 TTACGGCTAATGTGGGAAGTCTCTGAGGACATAAATATGCGCGTG..... 5922
992 alAspAlaTyrLeuAlaIleLysValAsnValGluLeuAlaIleLeuThr 1008
5923 .....AAGAAGAGAGATCCGAATTTGGA 5945
1009 GlyAspGlnArgGlnSerThrHisHisGlnGluArgGluSerGlnIleSe 1025
5946 AGGTCTCAACAGTGCTACCAAGTTAACTCCAGGCGTGAACATTTATCG 5995
1025 rSerLeuGlnSerAsnIleAlaGlnPheSerLysTyrAlaAspTyrTyrL 1042
5996 TCAGGGGAATCTAT..... 6009
1042 euAsnAlaThrHisArgGlnProArgArgLeuAlaAsnProIleLysVal 1058
6010 .....AAAGGGTTTGGTGGA 6026
1059 HisAlaGluArgGlnLeuGlyGlyAlaValLeuLysAlaAsnIleValPr 1075
6027 TGAGGTGCACATGATG.....CATCAAGCGTTACTA. 6057
1075 oAspLeuAlaMetValLeuValProAlaPheArgSerGlnSerLeuLeuT 1092
6058 ..CAACTAGCGCTTCTCCAACC.....GCCGCGTCGGAAGGCTCTTT 6099
1092 hrAspLeuGlyArgHisAlaMetThrTyrAlaGlyCysGlnGlyLeuThr 1108
6100 TTTGAGACATAAATCAGATA.....CCATTCAATAAAGAC 6134
1109 LeuAsnHisLeuThrIleIleLeuAspLysAspThrProLeuCysSerAs 1125
6135 GGAGAAGGTGTTTAGATGGATGTGCTGTTTGTTCCTCAAGAGGAA 6184
1125 pGluValLeuTyr.....ThrAlaPheSerArgAlaSerGluS 1138
6185 GCGTTGTATACACT.....TCTAAATCGTACAGGTGTCG 6219
1138 erIleThrPheValAsnThrHisSerAspAsnProAlaPheLeuAlaLys 1154
6220 TTAGATGTTTGTACTCTTCTCTCAATGACC.....GTAAGGGGAAC 6263
1155 LeuAspAlaThrProTyrLeuLysThrLeuIleSerTrpValArgGluAs 1171
6264 GGAAAGGTGTTACCTGAAAAGGTGCTTGAAGCGGTAAAGACAAACAGTAG 6313
||||: ||| |||||:

```

```

1171 pGluGlu.....AlaGlyAlaAspCysProAlaL 1181
6314 TAAGATCGCTGTCCAAAGG...CCAATTGGAACCACTGATGACGTAGCT 6360
1181 hrGluProLeuValLysAspValProThrLysThrHisIleProValAla 1197
6361 GAAATAAACGCTGAGCTGTACTTGTGCATGACCCAGTTGGGAGAAGTCGGA 6410
1198 AsnAspLysValGlnLeuGluGlyLysIleGluAlaMetGluAspLysAs 1214
6411 TATGAAGAGCGTCTTGAAGGAAAGAAAGAAACACACCATGATGATGACAG 6460
1214 pThrArgGluLeuTrpSerGlyGluGlyThrAsnLeuMetGlnThrG 1231
6461 TGCATGAAGCACAGGAGAAACATTCACGTGATGTGTGTTATTGTTAGGACG 6510
1231 lnAspProVal.....ValGlnLeuPheProHis 1240
6511 AAGAAAGCC...GATGACTCCCTATTTCATAAACAACCGCATATATCTGT 6557
1241 GlnGlnAlaLysAspGluAlaLeuPhe..... 1249
6558 TCGTTTTCGAGACACACACGCTCACTGTTTATCCCGCTCTGAGCTCAA 6607
1250 ..... 1250
6608 AGTTGGACGATAAGGTGCGCACATATATTAGCGCGCTCACCTCAATCA 6657
1250 ySile.....ThrIleGlyGluArgIleArgMetAlaThrProGluGln 1264
6658 GTATCCGACGCTTGTCTTCACACG 6681
1265 AsnAlaLysGlnLeuArgHisThr 1272
seq_name: pir2:JQ0096
seq_documentation_block:
hypoetical 176K protein - papaya mosaic virus
C:Species: papaya mosaic virus
C:date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C:Accession: JQ0096
R:Sit, T.L.; Abouhaidar, M.G.; Holy, S.
J. Gen. Virol. 70, 2325-2331, 1989
A:title: Nucleotide sequence of papaya mosaic virus RNA.
A:reference number: JQ0096; MUID:89381685
A:Accession: JQ0096
A:Molecule type: genomic RNA
A:Residues: 1-1547 <SIT>
A:Cross-references: GB:D13957; GB:D00580; NID:q222330; PIDN:BA003050.1; PID:q222331
A:Note: this protein shares two major domains of similarity with other potexvirus pro
C:Comment: The genome is a single-stranded, positive-sense RNA.
C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase
alignment_scores:
Quality: 404.00 Length: 798
Ratio: 1.101 Gaps: 35
Percent Similarity: 45.990 Percent Identity: 23.559
alignment_block:
US-09-301-906-3 x JQ0096 ..
Align seg 1/1 to: JQ0096 from: 1 to: 1547
4157 GTGAGTTGAGTGGACGCGCAAAAGGTTGTCCTCCCAAGAGCGTCTCGGT 4206
||| ||| :|: |||||: ||
405 ValThrLeuThrTyrProThrLysThrPheAspAsnArgGlyTrpValVa 421
4207 GAGGCGCGCACACAGAAGGGAAGAGCTGCAGATGCTAACACAGACACA 4256
|||: :|: :|: :|: ||| ||| :|: |||
421 lLysLeuGluAlaArgArgGlyTyrGluTrpPheAlaAspGluMetHisL 438
4257 AACCGCGGTCAACGAAG..... 4273

```

|||||:.....  
438 ysProLysGlyProGluLeuAsnLeuGluGluLysLysThrAspProAsp 454  
4274 ...GGACAGGAGGCCAGTACAG.....TCCAGTCTTGTGTAGTTC 4310  
|||||:.....  
455 AlaAlaSerTyrGluLysTyrLeuLysAlaLeuSerLeuLeuGlnLysG1 471  
4311 GCCACAGGCTGATATTCCAAAGGTCCACCAGTCCGAGGTACATGCTCAGA 4360  
|||||:.....  
471 uProGluValMetGluAlaLeuGluAlaThrAspGluProGlnA 488  
4361 AA...CAAGTGAACAAGATACCATTTGGCGACTCTTTCCGGCGCCAGC 4407  
:: |||||:.....  
488 rgProGluValLysGluGluGlnAlaGluAlaSerThrSerGlyArgAla 504  
4408 CCAATCGTGGATGAGAAACCCGCCCAAGTGTACGACTGGTGGTGAA 4457  
:|||||:.....  
505 GluGluLeuGlnGluAspProAlaThrLys..... 514  
4458 GATAATTGACAAGGCAAGCGCTCGCTCATGTGGCTGAGAAAAACAGG 4507  
515 .....LysGlyLys..... 517  
4508 TACAAGTCAGCAGCCCAAA...CAGAGGAGTTTGACGATCAATGAAGC 4554  
|||||:.....  
518 .....GluLeuProAsnProAsnArgAspLeu..... 526  
4555 AAGGCGGTAAACAGCTTTGCATGTTTGAACGTGTTCTCGCGGTGCA 4604  
||| |||||:.....  
527 .....LeuCysProCysGlyLeuH1 533  
4605 GCTGATGTCACAGGAGCCACTATCCACCAGGTTCTCAAC..... 4650  
||| |||:.....  
533 sLeuLysIleLysAsn.....AlaGluPheProGluLeuP 545  
4651 ..GCATTTACCTTTCTCGATAACTTGAAGGAGGAGTGCAGTCTTTTC 4698  
:|||||:.....  
545 roValLeuAspHisProAspHisLeuThrGlyArgLysAlaThrPhePhe 561  
4699 TCAAGCTGGGTGAGGGGTATACCTATATATGTTGGTACCATGTTTCATC 4748  
|||||:.....  
562 SerLysAspGlyLysProTyrSerTyrThrGlyGlySerHisAlaSerAr 578  
4749 AGGTTGGCTCTGCGCTAGAGGATATCTTAACGCCA.....ATTAAGT 4792  
|||||:.....  
578 gGlyTrpProAsnTrpLeuGluLysIleLeuAlaAlaIleGluIleLysG 595  
4793 ACCCA...AGCGTCTTCGACCACTCTTAGTCAGAGTACAGATGGGT 4839  
||| |||||:.....  
595 luProLeuProGluPheAsnGlnCysLeuValGlnGlnPheLysLeuGln 611  
4840 GGAGGCGTACCATTCCAGCTGATCAGCAGGAGTGTCTATCCATCAGATAA 4889  
:|||||:.....  
612 AlaAlaIleProPheHisArgAspAspGluProCysTyrProLysGlyHi 628  
4890 CCCTATCTGACGGTCAATCTCGTGGGAAGCAAACTCTCGATTAAGT 4939  
:|||||:.....  
628 sGlnValLeuThrIleAsnHisSerGlyLysLeuThrGlnIleAlaC 645  
4940 GCAGGAAGGTTGTAAGTTCATGGTTCATAAAGTACGTTCGGGTACACT 4989  
|||:|||||  
645 ysGlnLysGly.....LysAlaSerIleThrMetGlyPheGlyAspTyr 659  
4990 TTTCTTATGCTCGGTTTCCAAAGGAGCGCACTTGCATTCAGTAAACTC 5039  
:|||||  
660 TyrLeuSerProValGlyPheGlnGluSerHisLysHisAlaValSerAs 676  
5040 CATCGAAGGCGGCATCAGTTTCAGCTTCAGGCCAACT..... 5079  
:|||||:.....  
676 nThrThrGlyGlyArgValSerLeuThrPheArgCysThrValGlnGlnA 693  
5079 ..... 5079

693 snLysPheAsnAspAspGlySerMetGluAlaLeuAspAsnLeuProTrp 709  
5080 .....CGGCGCGTCTTTGGTGTAGGCAGGATGT 5108  
:|||||:.....  
710 LysAlaTrpIleProLysLeuGlnAsnLeuGlyPheGlnGlyArgGlnLe 726  
5109 GCAG.....TTACCGCGCGCTGTCGGATGAGAAGT 5140  
|||||  
726 uGlnTyrAspProAsnGlyAlaLeuIleSerProIleGluGluIleArg 743  
5141 CACCGAGTGTTCACAAACACCAACACAGAGC..... 5172  
|||  
743 er.....MetProLysCysLysProGluGlyValProGluValValTyr 757  
5173 .....CAAGGTGCTACCAAGACAATCACACAAATTCGGGGGCA 5213  
:|||||:.....  
758 LysThrLeuAspGlyLeuAlaArgAlaProThrProTyrSerProAsnPr 774  
5214 GGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATCT 5263  
:|||||:.....  
774 oIleArgAlaArgAlaTyrThrSerAspValLysAsnCysArgIleGlyA 791  
5264 CGATATGGTGCAGAACAAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGCT 5313  
:|||||:.....  
791 laLeuLeuArgGlnGlnGlyLys.....GluTrp..... 800  
5314 GATAATCCAGTATGAGTCTTGAACCTGACTACACCCCAATGACATTTGA 5363  
800 ..... 800  
5364 AGTGGTTAAACCGGGACCTCTGAAGATCGCGTGGTGGAGTACTTGAAGT 5413  
||| |||||:.....  
801 .....GlyCysArgPheAspAlaLeuValGluAlaGlyLys. 812  
5414 ATCTGCTATAGGCATTGAGAGACATACAGGCGTGTGCTTATGGCTAGA 5463  
813 .....Arg 813  
5464 AATATTCCGCTCACTACCGCCGAAGGTGT...CTGAAAGTACTTAATCA 5510  
:|||||:.....  
814 GluLeuAlaIleSerValIleHisGlyAlaGlyGlySerGlyLysSerG1 830  
5511 AGTTTATGAATCACTACCGGCTTTCACGTTTACAAGTCGGGCACAGATC 5560  
|||:|||||  
830 nAlaLeuGlnThrLeu..... 835  
5561 TCATTTTTTCACACACAGAGCGGTTCGCTGTGAGAGACCTACCGTAC 5610  
:|||||:.....  
836 .....IleLysAspAsnPro... 840  
5611 GTACTCATAGCTGAAAAAGGTATCTTTACCAAGGCAAGATGTCGACGC 5660  
841 .....GluLeuAspI1 844  
5661 GGTGTAGCTTTGGGCGACAATCTGTGTATGCGACAGATATACTGGTTT 5710  
:|||||:.....  
844 eThrValValLeuProThrAsn..... 851  
5711 TCCAGATGCCATTATTTGATAGGTGCTGAAAGTCTCGATCGATCGGC 5760  
:|||||:.....  
852 .....GluLeuArgLeuAspTrpLeuArgLysLeuProLysAlaPro 865  
5761 ATGGTGGGCAATCGTTTAACTTCGAAATATAGTGTATATGCTATGCTCC 5810  
:|||||:.....  
866 Gln.....GluLysPheLysThrPheGlu...LysAlaLeuLeuAlaPr 879  
5811 CCCAGTGGCGGTAAAGACGACGCTAGTGGACGAATTCGTTAAGTCAC 5860  
|||||  
879 oPro.....ThrProIleValIlePheAspAspTyrGlyLysLeuP 893  
5861 CCAAT.....AGCACGCCCACTT 5880  
|||  
893 roAlaGlyTyrValGluAlaPheCysLeuTyrPheSerThrValGlnLeu 909

```
5881 ACGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGTGAAGAAG 5930
      :::::
910 IleIleLeuThrGlyAspCys.....LysG1 918
5931 AGATCCGAAATTTGAAGTCTCACAGTGCTACACAGTTCACAGATTAACTCCAGG 5980
      :::::
918 nSerValHisHisGluSerAsnGluAsnAlaThrThrSerSer..... 932
5981 TGGTAAACTTTATCGTCACAGGGAATGTATAAA.....AGGTTTGTG 6024
      :::::
933 ..IleGluProLeuValLysGluAlaSerGluLeuCysArgTyrTyrIle 948
6025 GATGAGGTGCACATGATCATCAGGTTACTA...CAACTAGGCGTCTT 6071
      :::::
949 AsnAlaThrHisArgAsnLysLysAspLeuAlaAsnLysLeuGlyVal 965
6072 CGCA.....ACCGCGCGTCCGAA..... 6090
      :::::
965 rSerGluLysThrGlyLeuThrGluValThrHisGlyThrThrProIleP 982
6091 ..GCCTCTTTTGGGACACATAAATCAGATACCATTCATAAAGAGGAG 6138
      :::::
982 roGlyLeuHisMet.....LeuValProSerLeuThrLysLys 994
6139 AAGGTGTTAGG.....ATGGATTGTGC 6161
      :::::
995 GlnAlaPheSerGluMetGlyHisLysValSerThrTyrAlaGlyCysG1 1011
6162 TGTGTTTGTCCAAAGAGGAAGCGTTGTATACACTTCTAAATCGTACA 6211
      :::::
1011 nGlyIleThrAlaProLysIleGlnIleLeuLeuThrGluGluThrSerL 1028
6212 GGTGTCGTTAGATGTTGTACTTGTTCCTCAATCACCGTA 6255
      :::::
1028 euCysSerArgGluValLeuThrAlaLeuSerArgAlaVal 1042
```

seq\_name: pir2:S77910

seq\_documentation\_block:

hypothetical protein 1 - grapevine virus A

C:Species: grapevine virus A

C:Date: 18-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 20-Jun-2000

C:Accession: S77910

R:Minifra, A.

A:Reference number: S77910

A:Accession: S77910

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1707 <MIN>

A:Cross-references: EMBL:X75433; NID:g1405614; PIDN:CAA53182.1; PID:g1405615

C:Superfamily: grapevine trichovirus B hypothetical protein

alignment\_scores:

Quality: 394.00 Length: 1076

Ratio: 0.837 Gaps: 49

Percent Similarity: 43.773 Percent Identity: 21.654

alignment\_block:

US-09-301-906-3 x S77910 ..

Align seg 1/1 to: S77910 from: 1 to: 1707

3623 ATAATTGGTGGCGTGTCCCGAGTGTGTTGGCAGGGAAGAGTGTGAGC 3672

|||||

383 IleAlaGlySerLeuPheThrArgLeuPheAsnArgProGluTyrAspAl 399

|||||

3673 TCATTGCTTAC...TGCCTATTACCTAGGGGTATCTTTGATCATAAAGGG 3719

|||||

399 aArgCysLeuGluLysPheIleArg.....S 408

```
3720 CTGTGTTAAACGACACCATACCTCAACTTGCTTACGTCCACCGGTAGAGG 3769
      :::::
408 erCysGluThrThrGluIle.....His..ValGluA 418
3770 GGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTT..... 3813
      :::::
418 rGArg.....TyrMetGluGlyIleArgGlyAlaSerPheLysVal 432
3814 .....GACTATGACGAAGGTGCTGCTCCATCTGGGACTCAGCA 3851
      :::::
433 GlnAsnValMetAspTrpValGlu..... 440
3852 TGAAGCGGTTCCCGGTGACGATAACGATGGATGATCTTCTAGTGTCTCAA 3901
      :::::
441 .....AspAspSerAlaAsnAlaLeuSerGluValAsnP 452
3902 GCTATGATGTTGTCAAAATGTGCGCGACGTGGGATTAGCACCACCGG 3951
      :::::
452 heLeuAsp.....IleSerTrpAsnAsp 459
3952 GAAGTTACTGTGAAGAAGAGACCCATTCACCTCGAAGCGTCCAAATACAC 4001
      :::::
460 ArgValSer.....Glu.ProTyrGlyIleGluCysIleHisGlyG 473
4002 TTATGTCGAGGAAGAGGTGCGCGTCTGCGAGCTGTGCGGAAAGACAAAG 4051
      :::::
473 luGlySerArgIleArgValProLeuSerArgIleLeuArgAlaHisGlu 489
4052 GTGATCCGTCGGGTTCTGTTACCGCTGACCGTATGCGCTTTGTTGAAAGT 4101
      :::::
490 LeuIleAlaGlyValGlnThrAsp..... 497
4102 GTGAAAAAAGTGTGCGACGATGCTCTTTCCACCAACAGTCTAGTGGGAAAC 4151
      :::::
497 ..... 497
4152 GGTCTGTGAGGTGTGAGTGGGACGCAAGGTTGCTCCAGAAAGCGTCG 4201
      :::::
498 .....ValGluIleAsnPheProArgTyrValCysSerSerArgAlaLeu 512
4202 TCGGTGAGGCGCGACAC.....AAGAAAGGGAAGAGTGCAGATGTT 4245
      :::::
513 IleHisPheArgGlnTyrLeuIleLysLeuGlyArgPheSerPheMet... 528
4246 AACACAGCACAAACCGCGTCAACGAAGCGACAGGAGCCAGTACAGTC 4295
      :::::
529 .....GluSerArgAlaIleLys..AspIleGluAspIleGlnAl 541
4296 CAGTCTT.....GTGAGTTCCGCCACAGGCTGATATT..... 4326
      :::::
541 aGlyLeuGluGluGlyValIleThrGluGluAlaGluLeuArgLeuL 558
4327 .....CCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAA 4362
      :::::
558 euProThrThrLysProLysIleThr.....GluIleHisMetAsp... 571
4363 GAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACCCCAAT 4412
      :::::
571 ..... 571
4413 CGTCGATGAGAAACCGCCCAAGTGTTCAGACTCGTGGTGTGAAGATAA 4462
      :::::
572 .....AspAspThrProGly.....ThrSerGlyGluSerAspV 583
4463 TTGACAAAGGCAAGCGCTCGCTCATGTGGCTGAGAAAAAACAGGTACAA 4512
      :::::
583 alGluLysPheLysSerVal..... 589
4513 GTCGAGCAGCCCCAACACAGAGAGTTTGACGATCAATGAAGCAAGCGCG 4562
      :::::
589 ..... 589
4563 TAAACAGCTTTGCATGTTTAGAACGCTGTTCTCGGTGTGCAGCTGGATG 4612
```



590 ArgSerLeuCys.....ArgGluGluI 597  
4613 TGACAAACGAAGCACTATCGCCACCAAGTTCTCAACGCATTACCTTT 4662  
597 IefyrSerGlu..... 600  
4663 GTCGATAACTTGAAAGGAGGAGTGGGTCTTTTCTCAAGCTGGTGA 4712  
601 .....LysLeuLysGlyArgGluValAlaPheTyrSerHisSerLy 615  
4713 GGGGTATACCTATAATGGTGGTAGGCATCTTCATCAGGGTGGCTCGTG 4762  
615 sGluTyrLysTyrAsnGlySerHisArgSerLeuGlyTyrAspGluA 632  
4763 CCTAGAGGATATCTTAACGGCAATTAAGTACCCAAAGCCTCTTCGACCAC 4812  
632 laLeuAsnGluLeuThrGlnGluLeuGlyLeuAspAspSerTyrAspHis 648  
4813 TCTTTAGTCCAGAACTACAGATGGGTGGAGCGGTACCACTTCACGCTGA 4862  
649 CysLeuIleGlnArgTyrThrAlaGlyGlySerIleGlyPheHisAlaAs 665  
4863 TGACGAGGAGTGTCTCCATCAGATAACCTATCTTGACGGTCAATCTCG 4912  
665 pAspGluProCysTyrLeuProGlyGlySerValValThrValAsnLeuH 582  
4913 TGGGAAGCAAACTCTCGACTAAGTCAGGAAGGGTGGTAAAGGTCTGCA 4962  
682 isGlyAspAlaThrPheGluValLysGluAsnGlnSerGlyLysIleGlu 698  
4963 GTCATAAAGCTAGCTCGGTGCTACTATTCTTATGCTGCGGTTCCTCA 5012  
699 LysLysGluLeuHisAspGlyAspValTyrValMetGlyProGlyMetG1 715  
5013 AAGGAGCACTTGCATTCAGTAACTCCATCCAGCAAGGCGCATCAGTT 5062  
715 nGlnThrHisLysHisArgValThrSerHisThrAspGlyArgCysSerI 732  
5063 TCAGCTTCAGGCAACTCGGCCGCTCTTGGTGTAGCGAGGATGTTCAG 5112  
732 IefThrLeuArg.....AsnLysThrValAsp 740  
5113 TTAGCGGGCGGCTCGGATGAGAGTCACCAAGGTGTCCAAACCAAGCA 5162  
741 TyrGluAlaArgLysGlyAspGluAspSerGluTyrGluGluAspLysAl 757  
5163 ACCACAGACCAAGGTGCTTACCAGAACATCACACCAATCGGGGGCA 5212  
757 aGluLeuAspGluGlyIle..... 763  
5213 AGGCTCTATCTGAGGGAAGTGTAGGAAGTCAAGGGAGGTTCGACATAC 5262  
763 ..... 763  
5263 TCGATATGTCGACACAGATTACGTTAGGAAG..... 5295  
764 .....AspTyrLeuGlnLysAsnGlnLysMetCy 774  
5296 .....TGTGAGTGGCTCAGGCTGATATCCA....GTGA 5326  
774 sSerLeuLysAlaPheAlaAspHisMetGlnLeuSerThrProSerValI 791  
5327 TGGCTCTTGAACCTGACTACACCCCAATGACATTT...CAAGTGGTTAAA 5373  
791 leAlaIleValAsnGlyAlaSerProGlnThrLeuArgGluIleGluAsp 807  
5374 ACCGGAGCCTCTGAGATGCGCTGCTGGAGTACTTGAAGTATCTGCTAT 5423  
808 GlyGlyTyrSerLeuAlaThrLeuValAsnLeuSerLysAlaLeuAspPh 824  
5424 AGGCATTCGAGGACATACAGGCGCTGCTTATGCTAGAAATATATGCCG 5473  
.....

824 e.....ProIleAlaI 828  
5474 TCATACCGCCGAAGGTGTCTGAAATACCT.....AATCAAGTT 5514  
828 leHisGlyGluArgGlyTyrAlaGluThrProGlySerTyrArgArgLeu 844  
5515 TATGAATCACTACCGGGCTTTACAGTTTACAAAGTCGGGCACAGATCTCAT 5564  
845 HisLeuLysIleThrSerGlyHisValGlu..... 854  
5565 TTTTTCATCAACAACAGAGCGCTTGCCTGTGAGAGACCTTACCGTACGTAC 5614  
855 .....PropheGluG 858  
5615 TCATAGCTGAAAAGGTATCTTTTACCAAGGGCAAGATGTCGACGGGTG 5664  
858 lyValThrSerLysGlyGlyPheArgGlu.....Ala 868  
5665 GTAGCTTTGGGCGCAATCTGTTCGTATCGGACGATATACTGTTTTCCTCA 5714  
869 MetLeuLeuGlyAspGlyValGlyVal..... 877  
5715 CGATGCCATTAAATTCATAGTGCCTGAAAGTC.....GCTCGATGCG 5758  
878 .....GlyHisPheArgValAspLysAlaLysAlaA 888  
5759 GCATGTTGGGCGCAATCTTTTAAAGTCTTCCAAATATAAGTCTATAATGCT 5808  
888 spArgLeuAlaGlnSerPhe.....TyrAsn... 896  
5809 CCCCCAGGTGGGGTAAGACGACGAGCTTAGTGGACGAATTC..... 5850  
897 .....GlyAsnThrGlyValLeuLeuGlyLysTyrAsnLysG1 909  
5851 .....GTTAAGTCACCAATAGCACAGCCACCATTA 5881  
909 yLysMetHisThrGlyGluIleGluProLysGluValLeuThrAlaP 936  
5882 CGGCTAATGTGGGAAGT..... 5898  
926 heGlyPheAlaGlySerGlyLysSerHisTrpCysGlnThrIleLeuLys 942  
5899 .....TCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTT 5942  
943 HisCysSerValGluLysValLeuValIleSerProArgLysValLeuAr 959  
5943 GGAAGTCTCAACAGTCTACCACAGTTAACCTCCAGGTGGTAAACTTTA 5992  
959 gAspAspTrpValAlaLysIleSerLysLysHisArgValValThrPheG 976  
5993 TCGTCAGGGGAATG.....TATAAAAGGTTTGTGGTGGATGAG 6030  
976 luValAlaPheMetAspAspTyrGlyCysLysAspIleValIleAspGlu 992  
6031 GTGCACATGATGCATCAAGCTTACTACAACTAGAGCGTCTTCGCAACCGG 6080  
993 IleGlyLeuLeuProProGlyTyrIleAspLeuValIleAlaAlaHisG1 1009  
6081 CCGCTCGGAAGCGCTCTTTTTCGACACATAAATCAGATACCATTCATAA 6130  
1009 nProArgThrLeuValLeuLeuGlyAspProLeuGlnSerThrTyrHis 1026  
6131 ACAGGGAGAAAGGTGTTTAGGATCGATGTGCT.....GTTTGTGT 6171  
1026 erLysArgAspAsnValValLeuGluAlaSerGlnGluAspValPheAsn 1042  
6172 CCAAAGAAGAAAGCGTGTATACACTTCTAAATCGTACAGGTGTCCTGTT 6221  
1043 ArgValArgGlyLysLeuProTyrLeuCysTyrSerHisArgLeuProAr 1059  
6222 AGATGTTTCTACTTGTCTCTCAATGACCGTAAAGGGGAACGGAAGT 6271  
1059 gAsn...CysLysLeuPhe...GluIleGluCysMetGlyAlaGluSer... 1073



[illegible]

883 AlaLysLysLeuValAspSerPheGlnGlnGluGlyPheThrGlyVal..... 897

5508 TCAAGTTTATGAATCACTACCAGGGCTTTACAGCTTTTACAAAGTCGGGCACAG 5557

898.....CysLeuAsnLysPheGlnLysClnLysSerSerPheA 910

5558 ATCTCATTTTTCATTCAACACAAAGACGGCTTCGCTGTGAGAGACCTTACC 5607

910 snLeuVal.....AsnGlyLysGluLeuLeuAspValHis 921

5608 TAGCTACTCATAGCTCAAAAAGGTATCTTACCACAGGCAAGATGTCGA 5657

922 LeuThrLeu.....GlyPheAlaGlySerGlyLysSerPheTy 934

5658 CGCGTGGTAGCTTTGGGGCGCAACTCTGTCGTATGCGACGATATACTGG 5707

934 rProGlnCysValLeuLysAsnSer..... 942

5708 TTTTCCACGATGCCATTAAATTTGATAGTGGCACTGAAAGTCGCTCGATGC 5757

943.....HisTyraLaasnSerLeuValIleValProArgLysAlaLeuCys 957

5758 GGCATCGTGGCGGANTCGTTTAAAGTCCTTCGAATATAAGTCTCTATATGC 5807

958..SerAspTrpSerAsnLysValHisPro..... 966

5808 TCCCCAGGTGGCGGTAAAGACGACGACGCTTAGTGGACGAATTCGTTAAGT 5857

966..... 966

5858 CACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTCTTGAGGAC 5907

966..... 966

5908 ATAAATATCGCGGTGAAGAGAGAGATCCGAATTTGGNAGTCTCAACAG 5957

967.....AspValLysValThrPheGluSerA 976

5958 TGTACACAGTTAACTCCAGGGTGGTAACTTATCTCGTACGGGAATGT 6007

976 laPheArgGlnLysLysGlyTy..GlyLeuIleVal..... 988

6008 ATAAAAGGTTTTTGGTGGATGAGGTGCACATGCATCATCAAGGCTTACTA 6057

989.....IleAspGluIleGlyLeuLeuProProGlyTyIle 1000

6058 CAACATA.....GGGTCTTCCCAACCGCGCGTGGNAGGCGCTTTT 6101

1001 AspMetValHisGlyTyPheCysTyraAspSerLeu.....LeuValle 1015

6102 TGGAGACATAAATCAGTACCATTCAATAACAGGGAGNAGGTGTTTAGGA 6151

1015 uGlyAspProLeuGlnCysGluTyPheLysSerLysAlaAspHisPhe.... 1030

6152 TGGATTGTGCTGTTTTGTTCCAAAGAAAGAAAGCGTTGTA..... 6192

1031.....PheLeuGlyGlnGluSerValPheLysLysPhe 1042

6193.....TACACTTCTAAATCGTACAGGTGTCGGTTAGATGT 6227

1043 LysGlyHisCysAsnTyrlLeuTyrlLysSerHisArgLeuProArgAsn.. 1058

6228 TTCCTACTTGTTCCTCTANTGACCGTAAGGGGAACGGAAGTGTTCACC 6277

1059...GlnLysLeuPheGluIleGluCysAspGlyValGlu.....G 1071

6278 CTGAAAGGTCGTTAGCGCTAGGACAAACACAGTAGTAAGATCGCTGCC 6327

1071 LyGluGlyValThrPheAsnLysProArgAla..... 1081

6328 AAAAGGCCAATTGGAACCACTGTAGCTAGCTGAAATATAACCGTGCAGT 6377

1082.....LysAspLe 1084

```

6378 GTACTTGTGCATGACCCAGTTGGAGAACTCGGATATGAAGAGTGGTTGA 6427
      : ||||| :|||
1084 uThrLeuCysAlaSerGln.....A 1091
      :
6428 AGGGAAGGAAGAACACACAGTCAGTGCATGACAGTGAACACACAGGA 6477
      : ||| :||| ||| ||||| ||||| ||||| |||||
1091 rArgLysAspSerGluLysAspValAsnThrValGlyLysGlnGly 1107
      : ||| :||| ||| ||||| ||||| ||||| |||||
6478 AAACATTTCAGTGTGCTATTGTTTAGGACGAAGAACCGGATGACTC 6527
      : ||| :||| ||| ||||| ||||| ||||| |||||
1108 LeuSerAlaAsnArgValAsnIleuLeuAspLys.....AspTrpSe 1122
      : ||| :||| ||| ||||| ||||| ||||| |||||
6528 CCTA.....TTCACCT.....AAAC 6541
      : ||| :||| ||| ||||| ||||| ||||| |||||
1122 rLeuValAsnAspGluThrValIleValAlaPheThrArgAlaArgLysG 1139
      : ||| :||| ||| ||||| ||||| ||||| |||||
6542 ANCCGCATATA..... 6552
      : ||| :||| ||| ||||| ||||| ||||| |||||
1139 luIleAsnIleIleGlyAspAlaSerLeuValAsnAsnLeuLysArgSer 1155
      : ||| :||| ||| ||||| ||||| ||||| |||||
6553 .....CTTGTGTTGTGCGACACACACACG 6578
      : ||| :||| ||| ||||| ||||| ||||| |||||
1156 AlalysSerThrValLeuLysIleLeuGlyGlyGluArgValThrGI 1172
      : ||| :||| ||| ||||| ||||| ||||| |||||
6579 CTCACCTGTTATCCGCTCTGAGCTCAAAGTTGGACGATAAGGTC 6624
      : ||| :||| ||| ||||| ||||| ||||| |||||
1172 uGlyLeuIleLeuSerLeuIleArgLysLysLeuProAspValIle 1187
      : ||| :||| ||| ||||| ||||| ||||| |||||

```

seq\_name: pir2:S01865

seq\_documentation\_block:

```

genome polypeptide - tobacco rattle virus (strain SYM)
N:Alternate names: RNA replicase
N:Contains: 134K protein; 194K protein; RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: tobacco rattle virus, TRV
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 17-Mar-2000
C:Accession: S01865
R:Hamilton, W.D.O.; Boccardo, M.; Robinson, D.J.; Baulcombe, D.C.
J. Gen. Virol. 68, 2563-2575, 1987
A:Title: The complete nucleotide sequence of tobacco rattle virus RNA-1.
A:Reference number: S01865; MUID:88034943
A:Accession: S01865
A:Molecule type: genomic RNA
A:Residues: 1-1707 <HAM>
A:Cross-references: EMBL:X06172
C:Superfamily: cucumber mosaic virus RNA 1 protein
C:Keywords: nucleotidyltransferase
F:1-1707/Product: 194K protein #status predicted <MAT2>
F:1-1187/Product: 134K protein #status predicted <MAT1>
F:1188/Region: opal stop codon

```

alignment\_scores:

```

      Quality: 338.50      Length: 1858
      Ratio: 0.486        Gaps: 72
Percent Similarity: 37.513 Percent Identity: 17.976

```

alignment\_block:

US-09-301-906-3 x S01865 ..

Align seg 1/1 to: S01865 from: 1 to: 1707

```

1333 CACCAATGATTCGCGAAACCCTCA.....CATAG 1364
      ||| ||||| ||||| ||||| ||||| ||||| |||||
50 AspAspMetIleArgGluArgLysThrLysAspProValIleValHisGI 66
      : ||| :||| ||| ||||| ||||| ||||| |||||
1365 GTGCGATGTTCTTGAAGCGCGGGAAGGAGAGTGAAGAACTCT 1414
      : ||| :||| ||| ||||| ||||| ||||| |||||
66 u.....ValLeuSerGlnLysGluGlnAsnLysLeuMetGluIleT 80
      : ||| :||| ||| ||||| ||||| ||||| |||||
1415 TTCAGACGTTTCATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCA 1464
      : ||| :||| ||| ||||| ||||| ||||| |||||

```

```

80 yrProGluPheAsnIleValPheLysAspAspLysAsnMetValHisGly 96
      : ||||| ||| ||||| ||||| ||||| ||||| |||||
1465 TTCGTAATATGCATCGCGAGCTGTTTCAATGAATCTTTTCCAGGAGGTG 1514
      : ||||| ||| ||||| ||||| ||||| ||||| |||||
97 PheAlaAlaIaGluArgLysLeuGlnAlaLeuLeuLeuLeuAspArgVa 113
      : ||| :||| ||| ||||| ||||| ||||| |||||
1515 TGGTAATATGTGTC...TTCCTTCGATATTGGGGAGCTTCCAGTATCATG 1561
      : ||| :||| ||| ||||| ||||| ||||| |||||
113 lProAlaLeuGlnGluValAspAspIleGlyGlnTrpSerPheTrpV 130
      : ||| :||| ||| ||||| ||||| ||||| |||||
1562 TCRAAGCTGGCCATGTGAACCTGTCATGTATGCAATCCAGTCTTAGACGTT 1611
      : ||||| :||| ||| ||||| ||||| ||||| |||||
130 alThrArgGlyLysArgIleHisSerCysCysProAsnLeuAspIle 146
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1612 AAAGATGTGAAGCGGAGAACTCAATGAGATCTCTTTTCTTCCACAGCTGG 1661
      : ||||| :||| ||| ||||| ||||| ||||| |||||
147 ArgAspAspGlnArgGluIleSerArgGlnIlePheLeuThrAlaIleG 163
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1662 GGGAGATTCGTACGTGCCAGTGCACCTTCTTAACCTGAA..... 1698
      : ||||| :||| ||| ||||| ||||| ||||| |||||
163 yAspGlnAlaArgSerGlyLysArgGlnMetSerGluAsnGluLeuTrpM 180
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1699 .....GCGGCTTCAAAGTCTGTGTCTTACTGTT 1725
      : ||||| :||| ||| ||||| ||||| ||||| |||||
180 etTyrAspGlnPheArgLysAsnIleAlaAlaProAsnAlaValArgCys 196
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1736 AGTCGAGAATCGCAACTCGGATTTCTAGACCGGATCGGGTTTT..... 1770
      : ||||| :||| ||| ||||| ||||| ||||| |||||
197 AsnAsnThrTyrGlnGlyCysThrCysArg.....GlyPheSerAs 210
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1771 .....ATGGTGGATGTGTACG 1786
      : ||||| :||| ||| ||||| ||||| ||||| |||||
210 pGlyLysLysLysGlyAlaGlnTyrAlaIleAlaLeuHisSerLeuTyrA 227
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1787 ATATATCCCCCAGCAGTACAGAGCTATGGATAAGAGGGTCCGCTG 1836
      : ||||| :||| ||| ||||| ||||| ||||| |||||
227 spPheLeuLeuLysAspLeuMetAlaThrMetValGluLysLysThrLys 243
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1837 GTTTTCGACATAGTCTTATGTTTC...CCCGTGGAGTGTCTACGGTAA 1883
      : ||||| :||| ||| ||||| ||||| ||||| |||||
244 ValValHisAlaIleMetLeuPheAlaProGluSerMetLeuValAspGI 260
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1884 CGGTGAAGTTTACTTGAAGAAGCTGATACGTTGGTGAAGAGGAGGTG 1933
      : ||| :||| ||| ||||| ||||| ||||| |||||
260 uGly.....ProLeuProSerValAspGlyT 269
      : ||| :||| ||| ||||| ||||| ||||| |||||
1934 ATTACCTGGCCTACAATCTTGGTCAGTGTGGTGAGATGTATGAACATTC 1983
      : ||||| :||| ||| ||||| ||||| ||||| |||||
269 yrTyrMetLysLysAsn.....GlyLysIleTyr..... 278
      : ||||| :||| ||| ||||| ||||| ||||| |||||
1984 TTCCTAAACGTAAAGCGGCTTTTTC.....ACCTTTTCTTATGATACG 2024
      : ||||| :||| ||| ||||| ||||| ||||| |||||
279 .....PheGlyPheGluLysAspProSerPheSerTyrIleHI 291
      : ||||| :||| ||| ||||| ||||| ||||| |||||
2025 CACTTCGTCGGGAACGTGTTTAAGCTAGATGATGAGGATACCGTGTGTG 2074
      : ||||| :||| ||| ||||| ||||| ||||| |||||
291 sasTrpGlu.....GluTyrLysLysTyrLeuLeu. 301
      : ||||| :||| ||| ||||| ||||| ||||| |||||
2075 GTTACCATCATCTCACTATGTGTAGGGCTCAGAGTCACTTGGAACTGAG 2124
      : ||||| :||| ||| ||||| ||||| ||||| |||||
302 .....GlyLysPro 304
      : ||||| :||| ||| ||||| ||||| ||||| |||||
2125 GTTAGGTATAGTCTGTCCTCCGTCGTTGTCGGGCAATCGCTGCTGT 2174
      : ||||| :||| ||| ||||| ||||| ||||| |||||
305 ValSerTyrGln.....GlyAsnValPheTyrPh 314
      : ||||| :||| ||| ||||| ||||| ||||| |||||
2175 CATACCT...GTTCTAGCTGTTCTTAGTGTGCTCTTTAAGACAATAGTCC 2221
      : ||| :||| ||| ||||| ||||| ||||| |||||
314 eGluProTrpGlnValArgGlyAspThrMetLeuPheSer..... 327
      : ||||| :||| ||| ||||| ||||| ||||| |||||
2222 TCGATTCCGACTTTGTCACAGGATCTATTCTACCGCGCTCAACACTATA 2271
      : ||||| :||| ||| ||||| ||||| ||||| |||||
328 .....IleTyrArGIleAla..... 332
      : ||||| :||| ||| ||||| ||||| ||||| |||||

```

2272 GGCACATTCGAGAATAGAACGTTTGAGTATCGCGTTGGGCGGTCAGGTC 2321  
333 .....GlyValProArg.. 336  
2322 GCAAAAGACCATGTCATTACAGGAGTCGCGTTGTCCACAGCAAGGTTG 2371  
336 ..... 336  
2372 ATATTCTCTGATGATATATGTCGGGTTTAGTTCGCTGTTATGGCTCAG 2421  
336 ..... 336  
2422 GCGATTAAAGATAGGCGAAGAGTATTCGCTCCTATATAAAGC 2471  
337 .....ArgSerLeuSerSerGlnGluTyrTyrArg.. 346  
2472 CAGTGAGGGGAGTCTCCGCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAG 2521  
346 ..... 346  
2522 GCGATTGTTTTCGAACGAGCTCCGCTATAGCTAAGCAATGGTGCAC 2571  
346 ..... 346  
2572 GATAACTTCAACGTTTTGGAGAGCTTATGCTCTATGCCAGAGCGTTCAT 2621  
347 .....ArgIleTyrIle 350  
2622 CCGTAAAGTACCTGGTCTGTTGTTGTTACCAATTGCACCTTCGGAGCTT 2671  
350 eSerArgTTPGluAsnMetValValProIle. 361  
2672 CAGACAGGTGGAGCTCAGGGGTCCCTTTGATATTCGAAGGAGACCTTC 2721  
362 .....PheAspLeuValGluSerThr... 368  
2722 GGTAGAAACTGAAGAATAGTCGCTTGGCGCTTCTCTAGGCGCTATCGT 2771  
369 .....ArgGluLeuVal 372  
2772 GGAAGATTCAATTAAGGTCATGAAGCAATGAAGACAGAAGATGGAAAC 2821  
372 lLysLysAspLeuPheValGluLysGlnPheMet. 383  
2822 CCGTGCCAATTACTGAAGATTCGTATATGCGTTTCATAATGGG. 2865  
384 .....AspLysCysLeuAspTyrIleAlaArgLeuSer 394  
2866 .....AACGTTTCTACGTCCTACTGTACGAGGCGCTTCT 2903  
395 AspGlnGlnLeuThrIleSerAsnVal. 403  
2904 TGGCGGTTCCGAAGCGACGCTGTTTCGAGTGTTCCTAAGGGTTTGGTAG 2953  
404 .....LysSerTyrLeuSerSerAsnAsnTyrPvalLeuPheI 416  
2954 CTCGTGGGCTCGCAGCAAGCGCTTTCTCGGCATTACGTCGTTCTTTTCC 3003  
416 leAsnGlyAlaAlaValLys. 422  
3004 ACAGGTTCACTATTCTACACCGCGGTTTAACTGAGATGAAGGCTTGA 3053  
423 .....AsnLysGlnSerValAspSerArgAspLeuG1 433  
3054 TGCTCTGGTGGCACA. ....GAGAAATGCTATAAACTACCGGTGG 3094  
433 nLeuLeuAlaGlnThrLeuLeuValLysGluGlnValAlaAArgProVal. 449  
3095 GCATACTGGAGACGTGCGCGCTAGCTGTGAGCAAGGTCGTAGCTGGAAGC 3144  
450 .....MetArgGluLeuArgGluAlaIleLeuThrGluThr 461

3145 AAAGAATTTTGGAGTGAAGTTTCTTTAAATGACTTTCACCACCTTTCGTATT 3194  
462 LysProIle. ....ThrSerLeuThrAsp. 469  
3195 GCGGAATAAGTGCTTATCGGGATATTCTGGGTCTTTGGTCTCGGCC 3244  
470 .....ValLeuGlyLeuIleSerArgLysLeu. 478  
3245 CAATTCATGGAAGTATAGCGCGGAATTCGGCTAACGCTAGAGGTAC 3294  
479 .....TrpLys. ....GlnPhe 482  
3295 GCGGCGAGTAGTTACGAAACTTAAGCTCCTTAAGTTCAACAGCCGCGG 3344  
483 Ala. ....AsnLysIleAlaValGI 489  
3345 TCGTTTACGCGGTTTAACTCTAGCACAGTATCGGTGGATCTTTAGTCG 3394  
489 yGlyPheValGlyMetVal. ....GlyThrLeuIle. 499  
3395 TCGGAAGAGGTTTTCTCGTGGCGGTGACCTCACACTAGGCGGACCGTAGCT 3444  
500 .....GlyPheTyrProLysLysValLeuThr. 508  
3445 AAACGTCAAGTCCCTTAGCGTTGCTATCTTTTCTACCTCATACGCCAT 3494  
508 ..... 508  
3495 TTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACT 3544  
509 .....TyrPheLysAspThrProAsnGly. 516  
3545 TAATGTTTTTCTTGGTTTAGGACATGCTTGGGCGAGGCTAGCGCG 3594  
516 ..... 516  
3595 AATACTGGNAAGTTTGAGGCTTCTCCAATAATTGGTGCCTGTTCCCGA 3644  
517 .....ProGI 518  
3645 GGTGTTTGGCGAGGAAGAGTGTACGCTCATTTGTTACTGCCTATTACGC 3694  
518 uLeuCysTyrGluAsnSerHisLysThrLysValIleValPheLeuSerV 535  
3695 TA. ....GGGTATCTTTGATCATAAAGGGCTTCTTAAAC 3729  
535 alValTyrAlaIleGlyIleThrLeuMetArgArgAspIleArgAsp 551  
3730 GACACCATACCTCAACTTGTCTAGTCCCGACCGGTAGAG...GGAGGAA 3776  
552 GlyLeuValLysLysLeuCysAspMetPheAspIleLysArgGlyAlaH1 568  
3777 TGTGTACGAT. ....GAGACGCTTAGGTATTACCGGGACTTTTGACTATG 3820  
568 sValLeuAspValGluAsnProCysArgTyrTyr. .... 579  
3821 ACGAAGTCTGCTCCATCTCGGACTCAGCATGAAGCGTTCCCGGTGAC 3870  
580 .....Glu 580  
3871 GATAACGATGGATCCACTTCTAGTCTCTCAAGCTATGATGTTGTCACAA 3920  
581 IleAsnAspPhePheSerLeuTyrSerAla. .... 591  
3921 TGTGCGCGACGTGGGATTAGCACCAACGGGAAGTACTGTTGAAGAAG 3970  
592 .....SerGluSerGlyGluThr. .... 597  
3971 AGACCATTCACCTCGAAGCGTGCATATACACTTATGTCGAGGAAGAGTT 4020  
598 .....Val 598  
4021 GCCCCGCTGTCGACTGTGGCGGAAAGACAGGTGATCCGTCGGGTTCTGG 4070









3420 GACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCTTAGCGTTGC 3469  
564 e..... 564  
3470 TATCGTTTTTACCTCATACGCCATTTCGGGCTGCAGTATGTAGGCATT 3519  
564 ..... 564  
3520 TGGGCACATGCTCTCCACGGCACTTAATGTTTTTTCTTTGGTTTAGGGAC 3569  
565 :::: ||| ||| 574  
3570 ATTGCTTTGGGCGAGGCTAGCCGCAACTTGGGAAGTTTGGAGGCTTCT 3619  
574 sileAspGlyPro..... 578  
3620 CCAATAATTGGTGCCTGTTCOCGAGGTTTGGCGAGGAAGAGTGTC 3669  
578 ..... 578  
3670 AGCTCATTGTACTGCTATTACGCTAGGGGTATCTTTGATCATAGGGG 3719  
578 ..... 578  
3720 CTGTCTTAACGACACCATACCTCAACTTGTCTACGTCGCCACCGTAGAG 3769  
579 ..... SerProLeuIles 583  
3770 GGAGGAATGTGTACGATGAGAGCTTAGG.....TATTACCGGGACTTT 3813  
583 erAspProValtyrPheGlnSerLeuGluAspValtyrAlaGluTrp 599  
3814 GACTATGACGAAGGTCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCC 3863  
599 ..... 599  
3864 CGGTGACGATAACGATGATCCACTTCTAGTGTCTCAAGCTATGATGTTG 3913  
600 .....HisGlnGlyAsnAlaIleAspAlaSerAsnTyr..... 610  
3914 TCACAAATGTGGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGTT 3963  
610 ..... 610  
3964 GAAGAGACCCATTACCTCGAAGCTGCAATACACTTATGTCGAGGA 4013  
610 ..... 610  
4014 AGAGGTTGCCCGTCTGCAGCTGTGGCGAAGACAAGGTATCCGTCGG 4063  
610 ..... 610  
4064 GTTCTGTACCGCTGACGCTATGGCTTTTGTGAAGTGTGAAAAAGGT 4113  
611 .....AlaArgThrLeu 614  
4114 GTCGAGATGCTTTCACCAACAGTCTAGTGGGAAACGGCTCGTAGGT 4163  
615 LeuAspAspIleArgLysGlnLysGluGluSerLeuLysAlaLysAlaLys 631  
4164 TCAGGTGGACGCCAAAGGTTGCTCCCAAAAGCTGCTCGGTGAGGCC 4213  
631 sGluValGluAspAlaGlnLysLeuAsnArgAlaIleLeu..... 644  
4214 CGACACAAGAAAGGGAAGAGTGCAGATGGTAAACACAGCACAAACCGCG 4263  
644 ..... 644  
4264 GTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCC 4313  
644 ..... 644  
4314 ACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGTCACGAAG 4363

645 ..... 651  
4364 AAGTGAACAAGAAGTACCATTTGGCGACTGTTTCGGGCGCCACGCCAATC 4413  
651 luAlaHisProAsp..... 655  
4414 GTCGATCAGAAACCCGCCCAAGTTTACGACTCGTGGTGTGAAGATAAT 4463  
656 .....GlyGlyLysIleG 660  
4464 TGACAAGGCAAGCGCTCCCTCATGTGGCTGAGAAAAACAGGTACAAG 4513  
660 uGlyLeuGlyLeuSerSerGlnPheIleAlaLysIleProGluLeuAlaI 677  
4514 TCGAGCAGCCCAACACAGAGAGTTTGACGATCAATGAAGCAAGCGCGGT 4563  
677 leProThrProLysProLeuProGluPheGluLysAsnAlaGluThrGly 693  
4564 AAACAGCTTTTCATGTTTAGAACGTGTTCTCGCGGTGTGACGTGGATGT 4613  
694 GluIleLeu.....ArgIleAsnPr 700  
4614 GTACAAGGAGCGACTATCCACCAGGTTCTCAACGCCATTACCTTTG 4663  
700 oHisSerAspAlaIleLeu.....GluAlaIleAspTyrL 712  
4664 TCGATAACTTGAAGGAGGAGTGGCGTCTTTTCTCAAAGCTGGGTGAG 4713  
712 eulysSerThrSerAlaAsnSerIleIleThrLeuAsnLysLeuGly... 727  
4714 GGCTATACCTATAATGGTGTAGCCATGTTTCATCAGGGTGGCCTCGTGC 4763  
727 ..... 727  
4764 CTAGAGGATATCTTAACGGCAATTAACTACCAAGCGTCTTCGACCACT 4813  
728 .....AspHisC 730  
4814 GTTGTAGTCAGAAGTACAAGATGGGTGGAGCGGTACCATTCACGCTGT 4863  
730 ys.....GlnTrpThrLysGlyLeuAspValValTrpAlaGlyAsp 744  
4864 GAGCAGAGGTGCTATCCATCAGATAACCTTATCTTGACGGTCAATCTCGT 4913  
745 AspLys..... 746  
4914 GGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAAGTCATGG 4963  
746 ..... 746  
4964 TCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCTTCGGGTTTTCAA 5013  
746 ..... 746  
5014 AGACGCACTTGCATTTCACTAACTCCATCGACGAAGGCGCATAGTTT 5063  
746 ..... 746  
5064 GACGTTACGGCAACTCGCGCGCTCTTTGGTGTAGCAGGATGTTGCAGT 5113  
747 .....ArgArgAlaPhe..... 750  
5114 TAGCCGGCGCGTGTCCGGATGAGAAGTCACCAGGTGTTCCAACACAGCAA 5163  
751 .....IleProLysLysAsn 755  
5164 CCACAGACCAAGGTGCTTACAGACAATCACCACAAATTCGGGGGCAA 5213  
756 ThrTrpValGlyProThrAlaArgSerTyrProLeuAlaLysTyrGluAr 772  
5214 GGCTCTATCTGAGGGGAAGTGGTAGGGAAGTCAAGGGGAGGTTCGACATACT 5263  
:||||:||||

[illegible][illegible]

seq\_name: pir2:T45025

seq documentation block:

seq\_documentation\_prock:  
mucin MUC5B, tracheobronchial [imported] - human (fragment)

C;Species: Homo sapiens (man)

```
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
```

C;Accession: T45025

R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

J. Biol. Chem. 272, 3168-3178, 1997

**A;Title:** Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter

A;Reference number: Z22899; MUID:97166151

A;Accession: T45025

A; Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-3570 &lt;DES&gt;

A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503  
A;Experimental source: plasmids

A; Experimental source: placenta  
C: Genetics:

C;Genetics: A:Gene. MHC

A;Gene: MUC5B

## alignment\_scores:

Quality: 320.50 Length: 2469  
Ratio: 0.309 Gaps: 118  
Percent Similarity: 41.960 Percent Identity: 20.697

## alignment\_block:

US-09-301-906-3/rev x T45025

Align seg 1/1 to: T45025 from: 1 to: 3570

```
6430 CCTTCAACGACCTTCA.....TATCGACTTCTCCAACCTGGGTCAAG 6387
|||||.....|
1150 ProThrValThrSerSerLysAlaThrProSerSerProGlyThrAl 1166
|||||.....|
6386 CACAAGTACAGCTCAGCGTTTATTTTCAGCTAGCTCATCAGTGGTCCAAT 6337
|||||.....|
1166 aThrAla.....LeuProAlaLeuArgSer..... 1174
|||||.....|
6336 TGGCCTTTGGACAGGATCTTACTACTGTTGCTTCCCTACCGCTAACGA 6287
|||||.....|
1175 .....ThrAlaThr..... 1178
|||||.....|
6286 CCTTTTCAGGGTAACACTTTTCGTTCCCTTAGGTCATTGGAGACAAC 6237
|||||.....|
1179 ProThrAlaThrSerValThrProSerSerSerLeuGlyThrTh 1195
|||||.....|
6236 AAGTACGAACATCTAACGACACCTGTACGATTTAGAAAGTGATACAAC 6187
|||||.....|
1195 rTrpThrArg...LeuSerGlnThrThr..... 1204
|||||.....|
6186 GCTTTCCTCTTTGGNACAAAAACACCAATCCATCCTAACACCT... 6140
|||||.....|
1205 .....ProThrAlaThrMetSerThrAlaThrProSerSerThrProGlu 1219
|||||.....|
6139 .....TCTCCCTGTTTATGAATGGTATCTGA..... 6114
|||||.....|
1220 ThrAlaHisThrSerThrValLeuThrAlaThrAlaThrThrGlyAl 1236
|||||.....|
6113 .....TTTATGTCCTCCAAAAAGAGGCTTCCGACGCGCGCGTGG 6074
|||||.....|
1236 aThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr 1253
|||||.....|
6073 CGAAGACGCTAGT.....TGTAAGCTTGTATGC 6042
|||||.....|
1253 hrLysValProThrThrThrThrThrGlyPheThrAlaThrPro..... 1267
|||||.....|
6041 ATCATGTGCACCTCATCCACCAAAACCTTTATACATTCCCTGACGAT 5992
|||||.....|
1268 .....SerSerSerProGlyThrAlaLeuThrProProValTrpIl 1281
|||||.....|
5991 AAGTTTACACCCCTGGAGTTAACTGTGTAGTACTGTTGAGACCTTCCA 5942
|||||.....|
1281 eSerThrThrThrThrProThrProThrArgGlySerThrValThrProSerS 1298
|||||.....|
5941 AATTCGGATCTCTTCTTCCACGGCATATTATGTCCTCAGAACTTCCC 5892
|||||.....|
1298 erileProGlyThrThrHisThrAlaThrValLeuThrThrThrThr 1314
|||||.....|
5891 ACATTAGCCGTAATGTGGCTGTGCTATTGGGTGAGCTTAACGAATTCGTC 5842
|||||.....|
1315 ThrValAlaThr.....GlySerMetAlaThrProse 1325
|||||.....|
5841 CACTAACGTCGTCGTCTTATACCGCCACCTGGGGAGCATTTATAGCACTTAT 5792
|||||.....|
1325 rSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaT 1342
|||||.....|
5791 APTCGAAGACTTAACGATTCCGCCACCATCGCCGATCGAGCGACTTTC 5742
|||||.....|
1342 hrTrileThrAlaThrGlySerThrThrAsnProSerSerThrProGly 1358
|||||.....|
5741 AGTGCACCTATC....AATATATGCATCGTGGAAAAACCACTATATCGTC 5695
|||||.....|
```

```
1359 ThrArgProIleProProValLeuThrThrThrAlaThrThrProAlaAl 1375
5694 GCATACGAACAGATTGTGCGCCAAAGTACCACCGGTGCGCATCTTTGTC 5645
:.....|
1375 aThrSerSerThrValThrProSerSerAlaLeuGlyThrThrHisThrP 1392
5644 CTTTGGTAAAGATACCTTTTTCAGCTATGAGTACGTACGTACGTCTCTC 5595
||.....|
1392 ro.....ProValProAsnThrThrAlaThrThrHisGlyArgSerLeu 1406
5594 ACACGCAAGCGCTCTGTTGTGAATGAAAAATGAGATCTGTGCGCCGACTT 5545
:.....|
1407 Ser.....ProSer..... 1409
5544 GTAAACGTGAAAGCCCGGTAGTGAATTCATAAAGTTGATTAGGTACTTCA 5495
1409 ..... 1409
5494 GAACACCTTCGGCGGTAGTACGGCAATATTCTTAGCCATAAGCAAGCC 5445
:.....|
1410 ..SerProHisThrValArgThrAlaThrPThrSerAlaThrSerGlyThr 1425
5444 CTGTATGTCCTCTCAATGCCTATAGCCAGATCTTCAAGTACTCCAGCAC 5395
||.....|
1426 LeuGly.....ThrThrHisIleThrGluProSerTh 1436
5394 GGCACTTTCAGAGTCCCGGTTTAAACCACTTCAAAATGTCTATGGGGTGT 5345
:.....|
1436 rGlyThrSerHisThrProAlaAlaThrThrGlyThr..... 1449
5344 AGTCAGTTCAAGAGCCACTGCGATTATCAGCCCTGAGCCACTCACAC 5295
:.....|
1450 .....GlnHisSer... 1452
5294 TTCCTAACGTAATCTTGTTCGCACCATATGAGTATGTCGACCTCCCTT 5245
1453 .....ThrProAl 1455
5244 GACTTCCCTACCACTTCCCTCAGATAGAGCCTTG.....CCCGCGGATT 5201
||.....|
1455 aLeuSerSerProHisProSerSerArgThrThrGluSerProProSerP 1472
5200 TTGGTGTGATTGTTCTGGTAGCACCTTGGCTGTGCTGCTGGTTGGA 5151
||.....|
1472 roGlyThrThr..... 1475
5150 ACACCTGGTGACTTCTCATCCGACACGCGCGGCTAACTGCAACATCCT 5101
|||||.....|
1476 ThrProGly.....HisThrThrAlaThrSerArgThrThrAlaTh 1489
5100 GCTACACCA...AAGACGCGCGAGTTGCCCTGAACGTCAAACTGATGC 5054
|||||.....|
1489 rAlaThrProSerLysThrArg.....ThrSerThrLeuL 1501
5053 GCGCTTCGTGATGGAGTTTACTGAATGCAAGTGGCTCTTTGAAACCG 5004
|||||.....|
1501 euProSerSer.....Pro 1505
5003 CAAGGCATAAGAAATAGTACCCGGAAGCTACGTTTATGACCATGACCTT 4954
:.....|
1506 ThrSer.....AlaProIleThrThrValThrMetGlyCy 1518
4953 ACCACCTTCTGCACTTATGTCGAG.....AGTTTCCCTTCC 4916
:.....|
1518 s.GluProGlnCysAlaTrpSerGluTrpLeuAspTyrSerTyrPromet 1534
4915 CCACGAGATTGACCGTCAAGATAGGTTTATCTGATGATGATGACTCTCTCG 4866
||.....|
1535 ProGly.....ProSerGlyGlyAspPheAspThrTyrSerAsnIleAr 1549
4865 TCATCAGCGTGGAAATGTTAGCGCTCCACCATCTTGTACTTTCGCACTAA 4816
||.....|
1549 gAlaAlaGlyAlaValcysGluGlnPro..... 1559
```

4815 ACAGTGGTGAAGACGCTTGGTACTTAATTCCCGTTAAGATATCCTCTA 4766  
||||| |||:::|||||  
1560 .....LeuGly.LeuGluCysArg..... 1565  
4765 GGGCAGAGGCGACCTGATGAACATGGCTACCAACCATATAGTATAC 4716  
||| ::| |||:::|||||  
1566 .....AlaGlnAlaGlnProGlyValP 1573  
4715 CCCTCACCAGCTTTGAGAAAAGACCGCACTCCTCCTTTCAAGTTATC 4666  
||||| ::|::|  
1573 roLeuArgGluLeu.....GlyGlnValVal 1581  
4665 GACAAAGTAATGGTTGAGAACTGGTGGCGATAGTCGCTTCGTTGT 4616  
||| ::| ||| ::|  
1582 Glu.....CysSerLeuAspGly.LeuVal..... 1590  
4615 ACATATCAGCTGCACACCGGAGCAACAGCTTCTAAACATGCAAGCTGT 4566  
||| ::|::|::|::|::|::|  
1591 .....CysArgAsnArgGluGlnValGlyLysPheLysMetCys 1603  
4565 TTACGGCCTTCGCTTATGATGTCAACTCCTCTCTTTGGCTGCTC 4516  
::| ::| ::| ::|::|::|::|::|  
1604 Phe.....AsnIyrGluIleArgValPheCysCysAsnIyrGI 1616  
4515 GACTTGTACCTGTTTCTCAGCCACATGAGCGACGCGCTTGCCTTGT 4466  
::| |||::|::|::|::|::|::|  
1616 yHisCysProSer...ThrProAlaThrSerSerThrAlaThrProSerS 1632  
4465 CAATTATCTTCACACCAAGCTGTAACACTGGGGCGGGTTCVAVCG 4416  
|| ||| ::|::|::|::|  
1632 erThrProGlyThrThrTrpIleLeuThr..... 1641  
4415 ACGATTGGCGTGGCGCCGGAACAGTCCCAATGCTACTTCTGTTTCAC 4366  
::|::|::|::|::|::|::|::|  
1642 .....GluGlnThrThrAlaAlaThrThrAlaThrTh 1653  
4365 TTCTTCTGAGCATGACTCGGCTGGGTGACCTTTGGAATATCAGCCT 4316  
::| |||::|::|::|::|::|  
1653 rGlySerThrAlaIleProSerSer.....ThrProGlyThrAlaProp 1668  
4315 GTGGGAACTCACAAGCTGAGCTGCTACTGGCTCCTCTCGCTTCGCTG 4266  
::|::|::|::|::|::|::|::|  
1668 roProLysValLeuThrSerGlnAlaThrThrProThrAlaThrSerSer 1684  
4265 ACCGGGTTGTGCTGCTTACCATCTCCAGCT.....CTTCCCTTC 4222  
|||::|::|::|::|::|::|::|::|  
1685 LysAlaThrSerSerSerSerProArgThrAlaThrThrLeuProVal 1701  
4221 TTGTGTCGGCGCTCACCAGCAGCTTCTGGGAGCAACCTTTGCGCT 4172  
|||::|::|::|::|::|::|::|::|  
1701 uThrSerThrAlaThrLysSerThrAlaThrSerPheThrProIleProS 1718  
4171 CG.....ACCTCAACTCCAGACCGCTTTCCTCCA 4143  
|| |||::|::|::|::|::|  
1718 erSerThrLeuGlyThrThrGlyThrSerGlnAsnArgProHisPro 1734  
4142 CTAGACTGTTGS.....TGAAGACATCGTCGACACCTTTTTCACA.. 4101  
::|::|::|::|::|::|::|::|  
1735 MetAlaThrMetSerThrIleHisProSerSerThrProGluThrThrHi 1751  
4100 .....CTTTCACAAAAGCC.....ATAGCGTCAG 4076  
|||::|::|::|::|::|::|::|::|  
1751 sThrSerThrValLeuThrThrLysAlaThrThrThrArgAlaThrSers 1768  
4075 CGGTACCAAGACCGACGATCACT..... 4050  
::|::|::|::|::|::|::|::|  
1768 erMetSerThrProSerSerThrProGlyThrThrTrpIleLeuThrGlu 1784  
4049 .....TGCTTTCCGCGCACAGCTCA.....GACGGGCAACCTC 4015  
::|::|::|::|::|::|::|::|  
1785 LeuThrThrAlaAlaThrThrThrAlaAlaLeuProHisGlyThrProse 1801

4014 TTCCTCGACATAAGTGTATTGACGCTTCGAGGTGAATGGTCTCTCTCT 3965  
||| |||::|::|::|::|  
1801 rSer.....ThrProGlyThrThrTrpIleLeuThrG 1812  
3964 CACAGTAACCTCCCGTTGGTGTCTAATCCCAAG..... 3930  
::|::| |||::|::|::|::|::|  
1812 luProSerThrThrAlaThrValThrValProThrGlySerThrAlaThr 1828  
3929 ...TCGCGCACATTTGTGACAAACATCATAGCTTGACACACTA..... 3891  
||| ||| ::|::|::|::|::|::|::|::|  
1829 AlaSerSerThrArgAlaThrAlaThrLeuLysValLeuThrSerTh 1845  
3890 .....GAAGTGGATCCATCGTTATCTGT 3869  
::|::|::|::|::|::|::|::|  
1845 rAlaThrThrProThrValIleSerSerArgAlaThrProSer...SerS 1861  
3868 CACCGGGAACCGCTTCATGCTGAGTCCAGATGGACGACGACCTTCGTC 3819  
|||::|::|::|::|::|::|::|::|  
1861 erProGlyThrAlaThrAla.....LeuProAlaLeuArgSer 1873  
3818 TAGTCAAAAGTCCCGGTAAATACCTAAGCGTCTCATCGTACACATTCTCC 3769  
::|::|::|::|::|::|::|::|  
1874 ThrAlaThrThrPro.....ThrAlaThrSerValThrAlaIlePr 1887  
3768 CTCTACC.....GGTGGGACGTAAGCAA 3746  
|||::|::|::|::|::|::|::|  
1887 oSerSerSerLeuGlyThrAlaThrArgLeuSerGlnThrThrThrP 1904  
3745 GTTGAGGTATGCTGCTTAAGCAAGCCCTTATGATCAAGATACCCCT 3696  
::|::|::|::|::|::|::|::|  
1904 roThrAlaThrMetSerThrAlaThrPro.....SerSerThrPro 1917  
3695 AGCGTAATAGGCGAGTAACAATGAGCTGACACTCTTCCCTCGCCAAACA.. 3648  
::|::|::|::|::|::|::|::|  
1918 GluThrValHisThrSerThrValLeuThrThrThrAlaThrThrAr 1934  
3647 .....ACCTCGGGACAGCGCACCAAT 3626  
||| |||::|::|::|::|::|  
1934 gThrGlySerValAlaThrProSerSerThrProGlyThrAlaHis.... 1949  
3625 TATTGGGAAGCCTCCAACCTTCCAGATTCGCGCTAGCCTCGCCCA 3576  
||| |||::|::|::|::|::|::|::|  
1950 ..ThrThrLysValProThrThrThrGlyPheThrAlaThrPro 1965  
3575 AGCAATGTCCTTAACCAAGAAACAACTTAAGTCCGTGGGAAGAGCATG 3526  
|||::|::|::|::|::|::|::|  
1966 SerSerSerPro.....GlyThrAlaLeu 1973  
3525 TGCCCAA.....ATGCCTAACATACTGCAGCGCGAAATGGCGTATG 3485  
::|::|::|::|::|::|::|::|  
1973 uThrProProValTrpIleSerThrThrThrThrProThrThrArgGlyS 1990  
3484 AGGTAGAAAACGATAGCAACGCTAAGGAGCTTGAGTTTACGTAGGTC 3435  
::|::|::|::|::|::|::|::|  
1990 erThrValThrProSerSerIleProGlyThrThrHisThrAlaThrVal 2006  
3434 GGCCTAGTGACGGTCACC.....GCCGACGA 3409  
::|::|::|::|::|::|::|::|  
2007 LeuThrThrThrThrThrThrValAlaThrGlySerMetAlaThrProSe 2023  
3408 AAACCTCTTCGACGACTAAAGATCCACCGAT..... 3375  
::|::|::|::|::|::|::|::|  
2023 rSerSerThrGlnThrSerGlyThrProProSerLeuThrThrAlaThr 2040  
3374 .....ACTGTGCTAGAGTTAAACCGGTAAACCCACCGGG 3339  
||| |||::|::|::|::|::|::|::|  
2040 hrThrIleThrAlaThrGlySerThrThrAsnProSerSerThrProGly 2056  
3338 GCTTGTGAACCTTAACGAGCTTAGAGTTTCGTAACCTACTGCCCGCGTACT 3289  
::|::|::|::|::|::|::|::|  
2057 ThrThrPro.....IleProProVal 2064  
3288 TCTAGCGTTAGCCGAATTCGCGCTATATCTCATTCATTCGCGGCG 3239

[illegible]

2350 rThGlySerThrAlaThrProSerSerThrGlnGlyProProAlaGlyT 2367  
2509 AGAAGAGCTTGAAGACCCGGCGAGACTCCCTCAGTCGCTTTTATAAG 2460  
2367 hrProHisValSerThrAlaThrThrProThrValThrSerSerLys 2383  
2459 TTA.....TAGGAGCGATATCTTTCGGCCCTATC 2431  
2384 AlaThrProPheSerSerProGlyThrAlaThrAlaLeuProAlaLeuAr 2400  
2430 CTTAATCGCTGAGCATAACAGCAGCACAACTAAACCCACATATCATCAG 2381  
2400 gSerThrAlaThrThrProThrAlaThrSerPheThrAlaIleProSerS 2417  
2380 GAGAAATATCAACCTTGCTGG.....ACAACCGCA 2349  
2417 erSerLeuGlyThr...ThrTrpThrArgLeuSerGlnThrThrPro 2432  
2348 CTCCTGTAAATGACATGGGTCTTTTGGAGCTGACCGCCACCGGAT 2399  
2433 MetAlaThrMetSerThrAlaThrProSerSerThrProGluThrValH 2449  
2298 CTCAAAGCTTCTATTCTCGAATGTCCTTATAGTGTGAGCGCGTAGGAAT 2249  
2449 sThrSerThrValLeuThrThrThrAlaThrThrGlyAlaThrGlyS 2466  
2248 AGATCTGTGCAAAAGTCGGAATCGAGCACTATTGTCTTAAAGGACACA 2199  
2466 erValAlaThrProSerSerThrProGlyThrAlaHisThrLysVal 2482  
2198 CTGAACACGACTACACAGGT..... 2178  
2483 ProThrThrThrThrThrGlyPheThrValThrProSerSerSerProG 2499  
2177 .....ATCAACACAGCGATTTCGCCACGA 2153  
2499 yThrAlaArgThrProProValTrpIleSerThrThrThrProThr 2516  
2152 AGCAGCGGACCAACGACCTATACGTAACCTCAGTTCAGGTGACTTCTGA 2103  
2516 hrSerGlySerThr...ValThrProSerSerIleProGly..... 2528  
2102 GCCCTACATAGTAGATGATGGTAAACCACAGGTATCCCTCATACTC 2053  
2529 Thr...ThrHisThrProThrValLeuThrThrThrGlnProValAla 2544  
2052 TACGTTAAACACGCTCCCGGAGCAAGTCGGTACATAGAANAAGTGAAAA 2003  
2545 ThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrPr 2561  
2002 ACCCGCTTACGTTAGAGAAGGAATGTTTCATCATCTCACACACGTACCA 1953  
2561 oProSerLeuIle.....ThrThrAlaThrT 2570  
1952 ACATTGTAGGCGAGGTAAATCACTTCCCTCTTACCACACGTATCGAGTTC 1903  
2570 hrIleThrAlaThrGlySer.....ThrThrAsnProSerSe 2582  
1902 TTCCAAGTAAACTTCACGTTTACCGTACACACACTCCACGGGGAAACATA 1853  
2582 rThrProGlyThrThrProIlePro..... 2590  
1852 GAGCTATGTCGAAACACCGCACCTCTTATCCATAGCTCTGCTACC 1803  
2591 .....ProGluLeuThrThrThrAlaThrThr 2599  
1802 TGTGCGGGGATATATCGTACACATCCACCAATAAAACCGCATCGGCTCT 1753  
2600 .....ProAlaAlaThrSerSerThrValThrProSerSerAlaLe 2613  
1752 AGAATCGCAGTTCTCGGATTCGAGTTCAGATACAGTAAAGACACAGACTTGAAG 1703  
2613 uGlyThr.....ThrHisThrProProValProAsnThrThrAr 2626









4176 CAAAGGTTGCTCCAGAAAGCGTCGTGAGGCGCCGACACAAGAAA 4225  
622 .....  
4226 GGGGAAGAGCTGCAGATGTGTAACACACACAAAACCGCGGTCAACGAAGGC 4275  
624 erGluLysAlaSerAspGly..... 630  
4276 GACAGGAGCCAGTACAGTCTTGTGAGTTGCCACAGCTGATAT 4325  
631 .....  
4326 TCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAGTCAAAACAAG 4375  
641 uProSerIleLysGlySer.....MetAlaArgGlyG 652  
4376 AGTACCATTTGGAGCTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425  
652 luLeuGlnLeuAlaGlyLeuSerGly.....AspVal 662  
4426 CCCGCCCAAGTGTAGGACTCGTGTGTCGAGATAATTGACAAGGCCAA 4475  
663 ProGluSerSerTyrThr..... 668  
4476 GGCCGTCGCTCATGTGCTGAGAAAAAACAAGTACAAGTCGACGACGCCCA 4525  
669 .....ArgSerGluGluIleGluSerLeuG 677  
4526 AACAGAGAGTTTGACGTCATTAAGGCAAGCCGGTAAACAGCTTTGC 4575  
677 luGlnPheHisMetAlaThrAlaSerSerLeuIleHisLysGlnMetCys 693  
4576 ATGTTTAGAACGTGTCCTGCGGTGTGACGCTGGATGTGTACAAACGAAGC 4625  
694 SerIle.....ValTyrThrGlyPr 700  
4626 GACTATCGCCACCAAGCTCTCAACAGCATTTACCTTTGTGCTAACTTGA 4675  
700 oLeuLysValGlnGlnMetLysasn.....PheIleAspSerLeuV 714  
4676 AAGGGAGAGTGGCGTCTTTTCTCAAAGTGGGTGAGGGGTATACCTAT 4725  
714 aAlaSerLeuSerAla..... 719  
4726 AATGGTGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCTAGAGGATAT 4775  
720 .....AlaValSerAsnLe 724  
4776 CTTAACGGCAATTAAGTACCCAAAGCGTCTTCGACCACTGTTTGTAGTCAGA 4825  
724 uValLysIleLeuLysAspThrAlaAlaIleasp.....LeuGluThrA 739  
4826 AGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGC 4875  
739 rgGlnLysPheGly..... 743  
4876 TATCCATCAGATAACCCATATCTTGACGGTCAATCTCGTGGGGAAGCAAA 4925  
743 ..... 743  
4926 CTTCCTCGACTAAGTCAGGAAGGTGGTAGGTCATGTCATAAACGTAG 4975  
744 .....ValLeuAspValA 748  
4976 CTPCGGGTGAATTTCTTATGCCTTGC.....GGT 5007  
748 laSerLysargTrpLeuValLysProSerAlaLysAsnHisAlaTrpGly 764  
5008 TTTCAAGGACGCACTTGCAATTCAGTAAAC.....TCCATCGA 5045  
765 valValGluThrHisAlaArgLysTyrHisValAlaLeuLeuGluHisAs 781  
5046 CGAAGGGCCATCAGTTTGACGTTTCAGGGCAACTCGCGCGCTTTGGTG 5095

781 pGlu.....PheGlyI 785  
5096 TAGGCAGGATGTTGCAGTTAGCCGGGGCGTGTGGATGAGAAGTCACCA 5145  
785 le..... 785  
5146 GGTGTTCCAAACCAGCAACACAGAGCCCAAGGTGCTACCAGACAATCAC 5195  
785 ..... 785  
5196 ACCAAAATCGGGGCAAGGCTCTATCTGAGGAAGTGGTAGGGAAGTCA 5245  
785 ..... 785  
5246 AGGGAGGTGCACATATCGATATGTTGCGAACAAGATTACGTTAGGAAG 5295  
786 .....IleThrCysAsp..... 789  
5296 TGTGAGTGGCTCAGGCGTGATATCCAGTGATGGCTTTGAACCTGACTA 5345  
790 ...AsnTrpArg..... 793  
5346 CACCCCAATGACATTTCAAGTGGTTAAACCCGGGACCTCTGAAGTCGCG 5395  
794 .....ValAlaValSerSerGluSerV 801  
5396 TCGTGAGTACTTTGAAGTATCTGGCTATAGGCATTTGAGAGGACATACAG 5445  
801 alVal.....TyrSerAspMetAlaLysLeuArgThrLeuArg 813  
5446 GCCTTCTTATGGCTAGAAATATTCGCGTCACTACCGCCGAAAGGTGTCT 5495  
814 ArgLeuLeuLysAspGlyGluProHisValSerSerAlaLysValValLe 830  
5496 GAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTACGTTTACA 5545  
830 u..... 830  
5546 AGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTCGCGTGTG 5595  
830 ..... 830  
5596 AGAGACCTACCGTACGTACTCATAGCTAGCTGAAAGAGTATCTTTACCAAGGG 5645  
830 ..... 830  
5646 CAAAGATGTCGACGCGGTGTAGCTTTGGCGACAATCTGTTGCTATGCG 5695  
831 .....ValAspGlyVal..... 834  
5696 ACGATATACTGGTTTCCACGATGCCATTAAATTTGATAGTGCACCTGAAA 5745  
834 ..... 834  
5746 GTCGCTCGATGCGGCATGTTGGCGGAATCGTTTAAGTCTTCGAATATAA 5795  
834 ..... 834  
5796 GTGCTATAATGCTCCCGCAGGTGGGTAAAGACGACGACGCTTAGTGGACG 5845  
835 .....ProGlyCysGlyLys..... 839  
5846 AATTCGTTAAGTCACCCATAGCACAGCCACCATTACGGCTAATGTGGGA 5895  
840 .....ThrLysGluIleLeuSerArgValAsn 848  
5896 AGTTCTGAGGACATAAATATGGCG.....GTGAA 5924  
849 PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetII 865  
5925 GAAGAGAGATCCCAATTTGGAAGGTCTCAACAGTGTACACAGTAACT 5974  
..... 5974

```
865 eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882
5975 COAGGTGCTA...AACHTTATCGTC.....AGGGA 6003
      |||:||||| :|||:|||||
882 aArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898
      |||:||||| :|||:|||||
6004 ATGTATAAAGGTTTGGTGGATGAGTGCACATGATGATCAAGGCTT 6053
      |||:||||| :|||:|||||
899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyC 915
      |||:||||| :|||:|||||
6054 ACTACAACCTAGCGCTTCGCCAACCGCGCGTCGGAAGCGCTCTTTTGTG 6103
      |||:||||| :|||:|||||
915 sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932
      |||:||||| :|||:|||||
6104 CAGACATAAATCAGATACATTATCAACAGGAGGAGTGTGTTAGGATG 6153
      |||:||||| :|||:|||||
932 lYAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948
      |||:||||| :|||:|||||
6154 GATTGTGCTGTTTTTGTCCAAAGGAAGAAAGCGTTGTATACACTCTAA 6203
      |||:||||| :|||:|||||
949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965
      |||:||||| :|||:|||||
6204 ATCGTACAGGTGTCGTTAGATGTTTGTCTACTGTGTCCTCA..... 6246
      |||:||||| :|||:|||||
965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982
      |||:||||| :|||:|||||
6247 .....ATGACCGTAAAGCGGAAGCGGAAAGTGTACCTCAAAAG 6285
      |||:||||| :|||:|||||
982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998
      |||:||||| :|||:|||||
6286 GTCGTTAGCGGT.....AAGCAAAACAGTAGTAGTAAAGTCGCTGCCAA 6329
      |||:||||| :|||:|||||
999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl 1015
      |||:||||| :|||:|||||
6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACCTGACGTGT 6379
      |||:||||| :|||:|||||
1015 yys.....I 1017
6380 ACTTGTGTCATGACCCAGTTGGAGAGTCGGATATGAAGAGTTCGTTGAAG 6429
      |||:||||| :|||:|||||
1017 lCLeuThrPheThrGln.....SerAsp...LysGluAlaLeuLeu 1029
      |||:||||| :|||:|||||
6430 GGAAGAAGAAAGAACACACAGTGTACAGTGCATGATGAAGCACAGGAAA 6479
      |||:||||| :|||:|||||
1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyGl 1045
      |||:||||| :|||:|||||
6480 AACATTTCAGTGTGTTGTTTGTAGCAGCAAGAACCCGATGACTCCC 6529
      |||:||||| :|||:|||||
1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062
      |||:||||| :|||:|||||
6530 TATTCACTAAA...CAACCGCATATACATTGTTGTTGTCGAGACACACA 6576
      |||:||||| :|||:|||||
1062 lCileAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078
      |||:||||| :|||:|||||
6577 CGCTCACTGTTTATGCGCT.....CTGAGCTCAAAGTT 6611
      |||:||||| :|||:|||||
1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleI 1095
      |||:||||| :|||:|||||
6612 GCACCAT.....AAGGTCGCGCACATATATTAGCCAC 6642
      |||:||||| :|||:|||||
1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107
      |||:||||| :|||:|||||
```

seq\_name: pir2:s03701

seq\_documentation\_block:

141K protein - pea early browning virus

C:Species: pea early browning virus

C:Date: 19-Mar-1997 #sequence\_revision

C:Accession: S03701

R:MacFarlane, S.A.; Taylor, S.C.; King, D.I.; Hughes, G.; Davies, J.W.

Nucleic Acids Res. 17, 2245-2260, 1989

A:Title: Pea early browning virus RNA1 encodes four polypeptides including a putative zi

A:Reference number: S03701; MUID:89202038

A:Accession: S03701  
A>Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-1766 <MAC>  
A:Cross-references: EMBL:X14006; NID:g61019; PIDN:CAB37343.1; PID:g4456716  
C:Superfamily: cucumber mosaic virus RNA 1 protein

alignment\_scores:  
Quality: 306.00 Length: 1922  
Ratio: 0.406 Caps: 76  
Percent Similarity: 39.230 Percent Identity: 17.326  
alignment\_block:  
US-09-301-906-3 x S03701 ..  
Align seg 1/1 to: S03701 from: 1 to: 1766

```
1198 GTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTATCATCGTGGTTCGAG 1247
      |||:||||| :|||:|||||
7 lIeSerGlnLeuLeuAsnIleAspGluPheGlyAlaGluGlnArgGlyGl 23
      |||:||||| :|||:|||||
1248 TATCACCATAAGCTGGTATTAAAGGACCAATCCGCATTCGTTCCCATTT 1297
      |||:||||| :|||:|||||
23 nPheLeuAspLeuMetIleThrLysProAspSerGlnLeuGlyAlaMetM 40
      |||:||||| :|||:|||||
1298 TGGACACGAAATTTGCGCATATGTTTCTCAAAGGAGCGCAATGATTCGC 1347
      |||:||||| :|||:|||||
40 eGlnArgIleLeuThrAspLysIle.....AspAspSerLeuArg 53
      |||:||||| :|||:|||||
1348 GAAAAACCCCTCACATAGTGGCGATGTG.....TTTCTGAAGCCGCG 1388
      |||:||||| :|||:|||||
54 GluArgLysThrArgGluThrValIleHisGluLeuLeuSerGlnLy 70
      |||:||||| :|||:|||||
1389 GGAAGGAGAGACCTCAGGGAACCTCTTTCCAGAGCTTTCCGATACAGTCT 1438
      |||:||||| :|||:|||||
70 sAspGlnAsnLysLeuMetGluIleTyrProGluPheAsnIleValPheL 87
      |||:||||| :|||:|||||
1439 CCGACTCGTCAGGAGTAGTCACCCATTCGCTAATCGCATCGGAGCTGT 1488
      |||:||||| :|||:|||||
87 ysAspAspLysAsnMetValHisGlyPheAlaAlaAlaGluArgLysLeu 103
      |||:||||| :|||:|||||
1489 TTCATGGAATCTTTTCCAGGAGGTGTGTAATGTG...TGCTCTCTCGA 1535
      |||:||||| :|||:|||||
104 GlnAlaLeuLeuLeuMetAlaArgValProLysLeuGluProValAspAs 120
      |||:||||| :|||:|||||
1536 TATTGGGGGAGCTTCAGTATCATGTCAAAGCTGGCCATGTGAAGTGTCT 1585
      |||:||||| :|||:|||||
120 pIleGlyGlyGlnTrpSerPheTrpLeuSerArgGlyAspLysArgValH 137
      |||:||||| :|||:|||||
1586 ATGTATGCAATCCACTCTAGACAGTTAAAGATGTGAAGCGGAGAGATTC 1635
      |||:||||| :|||:|||||
137 lssSerSerCysProLeuIleAspMetArgAspLysGlnArgGluLeuGln 153
      |||:||||| :|||:|||||
1636 GAGATFCTCTTTCTTT.....TCCACAGCTGGGGGAGATTC 1670
      |||:||||| :|||:|||||
154 ArgGlnAsnPheLeuArgValPheArgAspAsnAlaThrThrSerGluSe 170
      |||:||||| :|||:|||||
1671 GTACGTGTCCAGTGACCTT.....CTAACTGAAGCGGCTTCAAAGTGTG 1714
      |||:||||| :|||:|||||
170 rArgIleSerAspAspGlnPheAspMetTyrAsnAlaPheLysGlyAspI 187
      |||:||||| :|||:|||||
1715 TCTCTTAC.....TGTAGTCGAGAATCCGACAACCTCGCAT 1749
      |||:||||| :|||:|||||
187 lEAspValAlaAsnPheValArgCysAsnAsnThrPheGlnAspCysAsn 203
      |||:||||| :|||:|||||
1750 TCTAGAGCCGATGCGGGTTTATTATGTTGGAT..... 1779
      |||:||||| :|||:|||||
204 CysArg.....GlyTyrArgAlaAspGlyThrArgIleGlyAlaTh 217
      |||:||||| :|||:|||||
1780 .....GTGTACATATATCCCGCAGCAGGAGTAG 1807
      |||:||||| :|||:|||||
217 rAsnAlaIleAlaLeuHisSerLeuTyrAspPheLysLeuAspValA 234
```



```

:::||||:||||:||||:||||:||||:
598 rgAsnCysGlyLeuLeuThrGlyTrpThrLysLeuLeu..... 610
3547 ATGTGTTTCTTGTAGGGACATTGCTGGGCGCAGGCTAGCGCAA 3596
610 ..... 610
3597 TACTTGGAGTTTGGAGGCTTCCAAATAATGGTGCCTGTTCCTCCGAGG 3646
611 .....AsnAsnThrHisAla.....HisV 617
3647 TTGTTTGGGAGGG.....AAGAGTGTCAAGCTCA 3675
|||:||||:||||:||||:||||:
617 aIGlyTrpGlnArgLeuPheLeuLysSerLysIleGluSerGluSerGlu 633
3676 TTGTTTACTGCTATTACCTAGGGGTATCTTTGATCATAGAGGGGCTTGT 3725
634 GlyArgLysProMetThr.....AspGluGluIleGluAlaAlaLe 647
3726 TAACGACACCATACCTCAACTTGCTTAGCTCCACCGGTAGAGGGAGGA 3775
|:||||:||||:||||:||||:
647 uAspAspIleMet.....G 652
3776 ATGTGTAGATGAGACGCTTAGGTATTACCGGAGCTTTGACTATGACGAA 3825
||||:||||:||||:||||:||||:
652 LuLeuAsnAspValAsnLeuGluAlaTyLys..... 662
3826 GGTGCTGCTCCATCTGGGACTCAGCATGAAGCGGTTCCTCCGGTGACGATAA 3875
662 ..... 662
3876 CGATGGATCCACTTCTAGTGTCTCA...AGCTATGATGTGTTCACA.... 3918
||||:||||:||||:||||:||||:
663 .....ThrArgThrValAsnLysGluPheAspIlePheThrThrT 676
3919 .....AATGTGGCGAGCTGGGATAGCACCAACGGGGAAGTTACT 3960
676 rPLeuAlaSerThrTyraThrGlyLeuAspSerGluLysGluLeuIle 692
3961 GGTGAAGAAGACACCATTCACCTCGAAGCGTCAATACACTATGTGCA 4010
692 ..... 692
4011 GGAAGAGTTGCCCGTCTGACGTGTGGCGAAGACAAGGTGATCCGT 4060
693 .....ThrAsnLeuLeuAlaThrAlaAlaValArgAsnLysLys..... 705
4061 CGGTTCTGGTACCCTGACGCTATGGCTTTGTTGAAAGTGTGAAAAA 4110
706 .....AlaLeuSerAspLysLeuAlaMetLeuIleAspValAspAsp 719
4111 GCTGTCGACGATCTTTCACCAACAGCTCTAGTGGGGAACGCTCGTGA 4160
720 SerValAsnSerPheLeuArgSerLeuSerAspThrAsp..... 732
4161 GGTGAGGTGGACGCAAGGGTTGCTCCAGAAAGCGTCTGCTGGTGAGG 4210
732 ..... 732
4211 CGCCGACACAAAGAGGGAGAGCTGCAGATGGTAACACA.....GCA 4254
733 .....AspAspSerThrAspValAla 739
4255 CAAACCGGGTCAACGAAGCGCACAGGAGCCAGTACAGTCCAGTCTTGT 4304
740 AspCysSerAlaSerValSerAspThrSerCysValSerValVa 756
4305 GAGTTCGCACAGGCTGATATCCAAAGGTCAACCCAGTCCGAGGTACATG 4354
756 lPheArgPro.....ThrValProGlyPheGlu...AspValAsnL 770
4355 CTCAGAAAGAGTGAACAAGAAGTACCATTGGCGACTGTTTCGGCGGCC 4404
:::||||:||||:||||:||||:

```

```

770 euArgLys.....GlyLys 774
4405 AGCCCAATCGTCGATGAGAAACCCGCCAAGCTTTACGACTCGTGGTGT 4454
||||:||||:||||:||||:||||:
775 ThrValIleValAspAsnAspAlaAlaGluSerSerSer..... 788
4455 GAAGATAATTGACAGGCGCAGCGCTCCTCATGTGGCTGAGAAAAAAC 4504
789 .....SerGluArgAsnA 793
4505 AGGTACAAGTCGACGAGCCCAACAGAGAGTTTGACGATCAATGAAGGC 4554
793 rg..... 793
4555 AAGGCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCTCTCGCGTGTGCA 4604
793 ..... 793
4605 GCTGATGTGTACAAAGCGACTATGCCACCAGGTTCTCAAACGCAT 4654
794 .....AsnHisPheAlaAsn...P 799
4655 TTACCTTTGTCGATAACTTTGAAAGGAGAGTGGCGTCTTTTCTCAAAG 4704
799 heGluValIleGluAsn.....CysArg 806
4705 CTGGGTGAGGGGTATACCTATATATGTTGATGCCATGTTTCATCAGGGTG 4754
807 PheGlyAspAlaProLysGluThrGlyAspPheSerValAspSerArg... 822
4755 GCCTCGTGCCTAGAGGATATCTTAACGGCAATTAAGTACCAACGCGTCT 4804
823 .....LeuGluPheIleHisTyLysLeu..... 829
4805 TCGACCACTGTTTACTG..... 4821
830 .....ArgCysLeuIleCysAlaGlnAsnAsnGluLeuLeuGlyLysTy 844
4822 CAGAAGTACAAGATGGTGGAGCGGTACCATTCACGCTGATGACGAGGA 4871
845 ArgAspTyrgluMetGly..... 850
4872 GTGCTATCCATCAGATAACCTTATCTTGACGGTCAATCTCTGTTGGGAAG 4921
850 ..... 850
4922 CAAACTTCTCGACTAAGTCAGGAAGGGTGGTAAG..... 4956
851 .....ValValArgProGlyGlyLysGlyTyProAspGlu 862
4957 GTCATGGTCAATAACGTAGCTTCGGGTGACTATTTCTTATATCCCT..... 5001
863 LeuGlyValPheAspLeuAlaLeuLysLysTrpIleIleLysProProse 879
5002 .TGCGGTTCCTCAAAGGACGACTTGCATTTCAGTAACTCCATCGACGAAG 5050
879 rCysSerTyraAsnLysAlaPheValProAspValSerAlaLysGluIng 896
5051 GCGCATCAGTTTGTGACGTTTCAGGCAACTCGCGCGCTCTTTGGTGTAGGC 5100
896 lyLys.....TrpIleGly 900
5101 AGGATGTTCCAGTTAGCCGGCGCGTGTCTGGATGAGAAGTCACCAGGTCT 5150
901 TyrLeuValAspAlaSerTrpGly..... 908
5151 TCCAAACCAAGCAACACAGAGCCAAGGTCTACCAGAACAAATCACACCAA 5200
909 .....L 909
5201 AATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGG 5250
909 ysGlnGlyIleAspAlaPheSer..... 916

```

```
5251 AGTCGACATACTCGATATGGTCCGAACAGATTACGTTAGGAAGTCTGA 5300
917      |||
5301 GTGGCTCAGGCTGATAATCCAGTGATGGCTCTTGAACCTGACGTACACCC 5350
918      |||||
919      .....TyrThra 920
5351 CAATGACATTGAAGTGGTTAAACCCGGGACCTCTGAAGATGCCGTCGTG 5400
920 snValAlaTrpLysAla..... 925
5401 GACTACTTGAAGTATCTGGCTATAGCATTTGACGAGGACATACAGGCGGTT 5450
925 ..... 925
5451 GCATTATGCTAGAAATATTGCCCTCACTACCGCCGAAGTGTCTCAAAG 5500
926 .....AspIleAlaIleValCysSerGln..... 933
5501 TACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCG 5550
933 ..... 933
5551 GGCACAGATCTCATTTTTCATTCAACACAGACGGCTTCGGTGTGAGAGA 5600
933 ..... 933
5601 CTTACCGTAGTACTACTAGCTGAAAAAGGTATCTTTACCAAGGGCAAAG 5650
933 ..... 933
5651 ATGTCGACGCGGTGGTAGCTTTGGCGACAATCTGTTCGTATCGACGAT 5700
934 .....ThrPheLeuCysAsnGlu 939
5701 ATACTCGTTTCCACGATGCCATTATTTGATAGGTGCACCTGAAATCGC 5750
940 ArgIleIleLeuLys.....AsnLeuAlaGly....LeuGluValVa 952
5751 T.....CGATCGGCATGGTGGCGCAATCGTTTAAAGTCTCTCGAATATA 5794
952 IProLeuArgCys.....LysPheL 959
5795 AGTGTCTATAATGCTCCCGCAGGTGGCGGTAAAGCAGCAGCTAGTGGAC 5844
959 ysLeuValAspGlyValProGlyCysGlyLysSerThrMetIleValAsn 975
5845 GAATTCGTTAAGTCAACCAATACGACAGCCACCATTTACGGCTAATGTGG 5894
976 .....ThrAlaAsnProValPheAspValValLeuSerThrCysLysG 990
5895 AAGTTCTGAGGACATA.....AATATGGCGGTGAAGAAGAGATCCCGA 5938
990 uAlaThrGluAspLeuLeuGluLysPheAlaAlaLysLysPheGlyIleA 1007
5939 ATTTGAAGGTCTCAACAGTGTACACAGTTAACTCCAGGTGTGTAAC 5988
1007 snLeuLys.....LysArgValLysThrValAspSerPheLeuMetHis 1021
5989 TTTATGCTCAGGGGAATGTATAAAAGGTTTTC.....GTGGATGAGGTGC 6034
1022 .CysSerAspGlyAsnCysValGlyAspLeuLeuHisPheAspGluAla 1038
6035 ACATGATGATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGCGCG 6084
1038 euMetAlaHisAlaGlyMetValPhePheCysAlaGlnIleAlaLysAla 1054
6085 TCGGAAGGCTCTTTTGGAGACATAAATCAGATACCATTCATATAACAG 6134
1055 LysLysValIleCysGlnGlyAspGlnLysGlnIleAlaTyrLysProAr 1071
```

```
6135 GGAGAAGGTGTTTAGGATGGATTGTCTGTTTGTTCCTCCAAAGAGGAAA 6184
1071 gValSerGlnLeuThrLeuArgPheThrSerLeuIleGlyArgPheAspG 1088
6185 GCCTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTAC 6234
1088 luValGluGluLysArgMetSerTyrArgCysProValaspValAlaLeu 1104
6235 TTCTTGTCTCAATGACCGGTAAAGGGGAACGGAAGTGTTCACCTGAAA 6284
1105 ThrLeu.....AspArgPheTyrThrGlyLy 1113
6285 GGTCCGTTAGCGGTAAAGCACAAACCACTAGTAGATCGCTGTCCAAAGGC 6334
1113 sValValThr.....LysAsnSerValLeuArgSerMetAspValLysA 1128
6335 CAATTGGAACCACTGATGACGTAGCTAGCTGAAATAAACGCTGACGTACTTG 6384
1128 rgIleGlySerLysGluGlnValGluMetGluHisGlyIleGlnTyrLeu 1144
6385 TGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTGCTTTGAAGGAAA 6434
1145 ThrPheLeuGlnSerGluLysLysAspIleAlaAsnLeuLeuCysGlnAr 1161
6435 AGCAAAAGAAACACCACTGATGATGACAGTCATGAAGCAGCACAGGAAAACAT 6484
1161 gLysValLysSerPheValAsnThrValHisGluAlaGlnGlyLysThrP 1178
6485 TCAGTGTATGTTGTTTGTAGGCAAGAAAGCGATGACTCCCTATTTC 6534
1178 heLysLysValArgLeuValArgLeuLysProThrAspAspValLeuAla 1194
6535 ACTAAACAACCGCATATCTTGTGTTGTCGAGACACACAGCTCACT 6584
1195 ArgGlyGlnGluTyrGlnIleValAlaLeuSerArgHisThrGlnSerLe 1211
6585 GGTTCATGCGCTCTGAGCTCAAGTTGGAGATAAGTCGGCGCACATATA 6634
1211 uValTyrGluThrValLys.....AspAspGluValSerAlaLeuI 1225
6635 TTAGCGACGCGTCA 6648
1225 leArgAspSerAla 1229
```

seq\_name: p1rl:WMTMPV

seq\_documentation\_block:

183K protein - pepper mild mottle virus (strain Spain)

N:Contains: 126K protein

C:Species: pepper mild mottle virus

C&gt;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 08-Apr-1994

C:Accession: JQ1312

R:Alonso, E.; Garcia-Luque, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra,

J. Gen. Virol. 72, 2875-2884, 1991

A:Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resist

A:Reference number: JQ1312; MUID:92113528

A:Accession: JQ1312

A:Molecule type: genomic RNA

A:Residues: 1-1611 &lt;ALO&gt;

A:Cross-references: GB:M81413

A:Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln

C:Comment: This protein may have RNA polymerase activity.

C:Superfamily: cucumber mosaic virus RNA 1 protein

F:1-1117/Product: 126K protein #status predicted &lt;PRO&gt;

alignment\_scores:

Quality: 297.00

Length: 854

Ratio: 0.754

Gaps: 40

Percent Similarity: 46.136

Percent Identity: 22.834

alignment\_block:

US-09-301-906-3 x WMTMPV

Align seg 1/1 to: WMTMPV from: 1 to: 1611

```

4522 CCCAACACAGAGTTTGACGATC..... 4545
|||||:|||||:|||||:|||||
366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuG1 382
|||||:|||||:|||||:|||||
4546 .AATGAAGCAAG.....GCCGGTAAACAGCTTTGCCATGTTTAGAAGCT 4588
|||||:|||||:|||||:|||||
382 nasnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAspP 399
|||||:|||||:|||||:|||||
4589 GTTCTCGCGGTGTCAGCTGGATGTGTCAAGAGCGACTATCCGCCACC 4638
|||||:|||||:|||||:|||||
399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu 415
|||||:|||||:|||||:|||||
4639 AGTTTCTCAACGCATTTACCTTTTCTGCATACATTGAAAGGGAGG..... 4683
|||||:|||||:|||||:|||||
416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValI1 432
|||||:|||||:|||||:|||||
4683 ..... 4683
432 eIleAsnGlyValThrAlaArgSerGluTrpAspValAspLysAlaLeuL 449
|||||:|||||:|||||:|||||
4684 .....AGTCGGGTCTTTTC.....TCAAAGCTGGT..... 4710
|||||:|||||:|||||:|||||
449 euGlnSerLeuSerMetThrPhePheLeuGlnThrLysLeuAlaMetLeu 465
|||||:|||||:|||||:|||||
4711 .....CAGCGGTATACCTATAATGGTGGTAGCCATGT 4742
|||||:|||||:|||||:|||||
466 LysAspAspLeuValValGlnLysPheGlnValHisSerLysSerLeuTh 482
|||||:|||||:|||||:|||||
4743 TTCATCAGGGTGG.....CCTC 4759
|||||:|||||:|||||:|||||
482 rGluTyrValTrpAspGluIleThrAlaAlaPheHisasnCysPheProT 499
|||||:|||||:|||||:|||||
4760 GTGCCCTAGAGGATATCTTAAG..... 4782
|||||:|||||:|||||:|||||
499 hrIleLysGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys 515
|||||:|||||:|||||:|||||
4783 .....GCATTTAAGTACCACCGCTC.....TTCAGCCACTGTTT 4817
|||||:|||||:|||||:|||||
516 AlaLeuGluIleLysValProAspLeuTyrValThrPheHisAspArgLe 532
|||||:|||||:|||||:|||||
4818 AGTCAGAGTACAAGATGGGTGGAGCGGTACCA.....TTCACAG 4858
|||||:|||||:|||||:|||||
532 uValLysGluTyrLysSerSerValGluMetProValLeuAspValLysL 549
|||||:|||||:|||||:|||||
4859 CTGATGACGAGGAGTGCTATCCATCAGATACCCCTATCTTGACGTCAT 4908
|||||:|||||:|||||:|||||
549 ysSerLeuGluGluAlaGluValMetTyrAsnAlaLeuSerGluIleSer 565
|||||:|||||:|||||:|||||
4909 CTCCTGGGGAAGGCAAC.....TTCCTCCTAAGTG 4940
|||||:|||||:|||||:|||||
566 IleLeuLysAspSerAspLysPheAspValAspValPheSerArgMetCy 582
|||||:|||||:|||||:|||||
4941 CAGGAAGGTGT.....AAGGTCATGGTCATAA 4969
|||||:|||||:|||||:|||||
582 sasnthrLeuGlyValAspProLeuValAlaAlaLysValMetValAlaV 599
|||||:|||||:|||||:|||||
4970 AGCTAGCTCGGGTGACTATTTTCTTATGCCCTTGCGGTTTCAAAGGACG 5019
|||||:|||||:|||||:|||||
599 alValSer..... 601
|||||:|||||:|||||:|||||
5020 CACTTGCATTAGTAACTCCATCCACGAAGCGGCATCAGTTTACGTT 5069
|||||:|||||:|||||:|||||
602 .....AsnGluSerGlyLeuThrLeuThrPh 610
|||||:|||||:|||||:|||||
5070 CAGGCAACTCGCGCGTCTTTGGTGTAGGCAGGATGTTGACGTTAGCCG 5119
|||||:|||||:|||||:|||||
610 eGluArgProThrGluAla....AsnValAlaLeuAlaLeuGln..... 623
|||||:|||||:|||||:|||||
5120 GCGGCGTCTCGGATGAGAAGTCACCGAGTGTTCACAAACGACGACACAG 5169
|||||:|||||:|||||:|||||

```

```

624 .....ProThrIleThrSerLys 629
5170 AGCCAAGGTGTACCAACAATACACAAATCGGGGGCAAGGCTCT 5219
|||||:|||||:|||||:|||||
630 GluGluGlySerLeuLysIleValSerSerAspValGly..... 642
|||||:|||||:|||||:|||||
5220 ATCTCAGGGAAGTGTAGGAAGTCAAGGGAGTCCGACATACTCGATA. 5268
|||||:|||||:|||||:|||||
643 ....GluSerSerIleLysGluValValArgLysSerGluIleSerMetL 658
|||||:|||||:|||||:|||||
5269 .....TGGTCGCAACAAGATTACGTTAGGAAGTGTGAG 5301
|||||:|||||:|||||:|||||
658 euGlyLeuThrGlyAsnThrValSerAspGluPheGlnArgSerThrG1 674
|||||:|||||:|||||:|||||
5302 TGGCTCAGGGCTGATACTCAGTGTGCTCTTGAACCTGACTACACCCC 5351
|||||:|||||:|||||:|||||
675 .....IleGluSerLeuGln..... 679
5352 AATGACATTTGAAGTGTAAACCGGGACC.....TCTGAAGATG 5392
|||||:|||||:|||||:|||||
680 ....GlnPheHisMetValSerThrGluThrIleIleArgLysGlnMeth 695
|||||:|||||:|||||:|||||
5393 CGGTCTGGAGTACTTGAAGTATCTGGGTATAGGCATTGAGAGGACATAC 5442
|||||:|||||:|||||:|||||
695 isAlaMetValTyrThrGlyProLeuLysValGlnGlnCysLysAsnTyr 711
|||||:|||||:|||||:|||||
5443 ....AGGGCGTTCCTTATGGCTAGAAATATTGCCGTCACCTACCCCGAAG 5489
|||||:|||||:|||||:|||||
712 LeuAspSerLeuValAlaSerLeuSerAlaAlaValSerAsnLeuLysL 728
|||||:|||||:|||||:|||||
5490 TGTTCTGAAAGTACCTAATACTCAAGTTTATGAATCACTACCGGGCTTTCAG 5539
|||||:|||||:|||||:|||||
728 sIleIleLys..... 731
5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATCAACACAGACGCTTG 5589
|||||:|||||:|||||:|||||
732 .....AspThrAlaAlaIleAspLeuGluThrLysGluLysPhe 744
|||||:|||||:|||||:|||||
5590 CGTGTACAGACCTACCTACGTACTACTAGCTGAAAGAGGTATC..... 5634
|||||:|||||:|||||:|||||
745 GlyVal.....TyrAspValCysLeuLysLysTrpLeuVally 757
|||||:|||||:|||||:|||||
5635 ....TTTACCAGGCAAGATGTCGACGCGGTGGTAGCTTTGGCGACA 5680
|||||:|||||:|||||:|||||
757 sProLeuSerLysGlyHisAlaTrpGlyValValMet.....Asps 771
|||||:|||||:|||||:|||||
5681 ATCTCTTCTGATGCGACGATATACTGGTTTCCAGATGCCATTAATTG 5730
|||||:|||||:|||||:|||||
771 exAspTyrLysCysPheValAlaLeuLeuThrTyrAspGlyGluAsnIle 787
|||||:|||||:|||||:|||||
5731 ATAGTGTCACTCAAAGTCGCTCGATCGGC.....ATGGTGGG 5768
|||||:|||||:|||||:|||||
788 Val.....CysGlyGluThrTrpArgArgValAl 797
|||||:|||||:|||||:|||||
5769 CGAATCGTTTAAGTCCTTCAATATAAAGTGCTAT..... 5802
|||||:|||||:|||||:|||||
797 aValSerSerGluSerLeuValTyrSerAspMetGlyLysIleArgAlaI 814
|||||:|||||:|||||:|||||
5803 .....AATGTCCCCAGGTGGCGGTAAAGACGAG 5832
|||||:|||||:|||||:|||||
814 leArgSerValLeuLysAspGlyGluProHisIleSerSerAlaLysVal 830
|||||:|||||:|||||:|||||
5833 AGCTTAGTCGACGAATTCGTTAAGTCACCAATAGCACAGCCACCATTAC 5882
|||||:|||||:|||||:|||||
831 ThrLeuValAspGly...ValProGlyCysGlyLysThrLysGluIleLe 846
|||||:|||||:|||||:|||||
5883 GGCTAATGTGGGAAGTCTCTGAGGACATAAATATGGCGTGAAGAAG.... 5928
|||||:|||||:|||||:|||||
846 uSerArgValAsnPheAspGluAspLeuValLeuValProGlyLysGlnA 863
|||||:|||||:|||||:|||||
5929 .....AGAGATCCGAATTTGGAAGGTCTCAACAGTCT 5961
|||||:|||||:|||||:|||||
863 laAlaGluMetIleArgArgAlaAsnSerSerGlyLeuIleValAla 879
|||||:|||||:|||||:|||||

```



205 rIleTyrAspIleProTyrSerSerIleGlyProAlaLeuHisArgLysA 222  
1829 GTGCGCTGTTTCCACATAGCTCTATGTTCCCGGTGGAGTTGTTGAC 1878  
222 snValArgValCysTyrAlaAlaPheHisPheSerGluAlaLeuLeuLeu 238  
1879 GGTAAAGGTGAAGTTTACTTGAAGAACTCGATACGTTGGTCAAGAGGA 1928  
239 GlySerProValGlyAsnLeuAsnSerIleGlyAlaGlnPheArgValAs 255  
1929 AGTTCATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATG...TATG 1975  
255 pGlyAspAspValHisPheLeuPheSerGluGluSerThrLeuHisTyrT 272  
1976 AACATTCTCTCTAACCTAAGCGGTTTTCACCTTTCTTATGATACGC 2025  
272 hrHisSerLeuGluAsnIleLeuLeuIleValMetArgThrTyrPhePro 288  
2026 ACTTCGTCGGGAAGCTGTTAAAGCTAGATATGAGGATACCGTTGTGG 2075  
289 AlaAspAspArgPheValTyrIleLysGluPheMetValLysArgValAs 305  
2076 TTACCATCATCTCACTATGTTAGGGCT.....CAGAAST 2110  
305 pThrPhePhePheArgLeuValArgAlaAspThrHisMetLeuHisLys 322  
2111 CACCTGGAACCTGAGTTACGTATAGTGC.....TTG 2142  
322 crValGlyHisTyrSerLysTrpLysSerGluTyrPheAlaLeuAsnThr 338  
2143 CTCCGCTCTTCGTCGGCAATCGCTG..... 2169  
339 ProProIlePheGlnAspLysAlaThrPheSerValTrpPheProGluAl 355  
2170 .....GTGTTTCACTGTT.....GTAGCTGGTTCTA 2197  
355 alyLysValLeuIleProLysPheGluLeuSerArgPheLeuSerGlyA 372  
2198 GTGTCCTCTTAAGACATAGTCTCGATTGGACTTTGTGCAGAGGATC 2247  
372 snValLysIleSerArgMetLeuValAspAlaAspPheValHisThrIle 388  
2248 TATTCTCCTACCGCTCAACATATAGGACATTCGAGAATAGA.....AC 2291  
389 .....IleAsnHisIleSerThrTyrAspAsnLysAlaLeuVa 401  
2292 GTTTCAGTATCCGTTGGGCGGTGAGTCCGAAAGACCATGTCATTA 2341  
401 lTrpLysAsnValGlnSerPheValGluSerIleArgSerArgValIleV 418  
2342 CAGGAGTCGCTGTGTCACACGAGGTTGATATTCTCTCGATGATATG 2391  
418 aIAsnGlyValSerValLysSerGluTrpAsnValProValAspGlnLeu 434  
2392 TCGGTTTATGTCGCTCTATGGCTCAGCGATTAAAGGATAGGGCGAA 2441  
435 ThrAspIleSerPheSerIlePhe...ProLeuValLysValArgLysVa 450  
2442 CAGTATTCGC.....TCTATAACTTTTATAAAGCCAGTGAGGGAGTC 2485  
450 lGlnIleGluLeuMetSerAspLysValIleGluAlaArgGlyLeuL 467  
2486 TC.....GCCGGGTCTTCAGCTCTTCTTCAGACCGTAGCGCAT 2526  
467 euArgArgPheAlaAspSerLeuLysSerAlaValGluGlyLeuGlyAsp 483  
2527 TGTTTTTCGAACGAGTCTCCGCTCTATGCTAAGGCAATGGTGCACGATA 2576  
484 Cys.....ValTyr..... 486  
2577 CTTCAACGTTTTTGGAGACGCTTATGCTCTATGCCAGACGCTTCATCGTA 2626

486 ..... 486  
2627 AAGTACCTGGGCTCTGTGTGTGTACCATTTGCACCTTCGGAGCTTCAGAC 2676  
487 .....Asp 487  
2677 AGTTTGGAGCTCAGGGGTGCCCTTTGATATTTTGAAGAGACCTTCGGTAG 2726  
488 AlaLeuValGlnThrGlyTrpPheAspThrSer..... 498  
2727 GAACTGAAGAAATAGTCGCTTCGGCTCTTCTTAGGGCTATCGTGGAG 2776  
499 .....SerA 500  
2777 ATTCAATTAAGTTCATGAAGCAATGAACACAGAAAGATGGAACCCCTG 2826  
500 spGluLeuLysValLeu.....LeuProGlu 508  
2827 CCAATTACTGAAGATTCTCTATATCGTTTCATATGGGGAAGCTTCTAA 2876  
509 ProPheMetThrPheSerAspTyrLeu..... 517  
2877 CGTCCACTGTACGAGGCGAGGCTCTTCTTGGCGGTTTCAAAGCAGCGTGS 2926  
518 .....GluGlyMetTyrGluAlaAspAlaLysIleGluA 529  
2927 TTTCCAGTGTCTTAAGGCTTTGGTAGCTCGTGGGGCTGCGACGAGGCC 2976  
529 rgGluSerValSerGluLeuAlaSerGlyAspLeuPheLysLys 545  
2977 TTTCTGCGATTACGTCGTCTTTTCCACAGGTTCACTATTCTACGACCG 3026  
546 IleAspGluIleArgAsnAsnTyrSer..... 554  
3027 CGTTTAACTGAAGATGAAGGCTTGATGCTCTGTGCGCACAGAAATG 3076  
555 .GlyValGluPheAsp..... 559  
3077 CTATAAACTCACCGGTGGCATACTGGAGACGTCGCGGTAGCTGTGAGC 3126  
559 ..... 559  
3127 AAGTCGTAGCTGGAACGAAGAATTTTGGAGTGAAGTTTCTTAAATCA 3176  
560 .....ValGluLysPheGlnGluPheCysLysGluLeuAsnVal..... 572  
3177 CTTCAACACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGG 3226  
573 .....AsnProMetLeuIleGly..... 578  
3227 CGTCTTTGGTGGCGGCCCAATTGTCATGGAAGTATAGCGCGGAATTGCG 3276  
578 ..... 578  
3277 GCTAACGCTAGAAGGTACCGGGCAGTAGTTACGAAACTCTAAGCTCGTT 3326  
579 .....HisValIleGluAlaIle 584  
3327 AAGTTCACAAGCCGCGGTGTGTTACGCGGTTTAACTCTAGCACAGTAT 3376  
584 ePheSerGlnLysAlaGly..... 590  
3377 CCGGTGGATCTTTAGTCGTGCGAAGAGGTTTTCGTCGCGGTGACCGTC 3426  
591 .....ValThrVal 593  
3427 ACTAGGCGGACCGTAGCTAAACGCTCAAGTCCCTTAGCGTTCTATCGTT 3476  
594 Thr..... 594  
3477 TTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCAC 3526  
594 ..... 594



3527 ATGCTCTTCCACGGCACTTAATGTTTTTCTTGGTTTAGGCACATTGCTT 3576  
|||||  
595 .....GlyLeuGlyThrLeu... 599  
3577 GGGCGAGGGCTAGCCGAATACTTGGAGTTTGGAGGCTTCTCCAATAA 3626  
599 ..... 599  
3627 TTGGTCGGTGTTCGGAGGTTGTTTGGCAGGGAAGAGTGTACAGCTCAT 3676  
|||||  
600 .....SerProGluMet.....GlyAlaSerValAla... 608  
3677 TGTACTGCCTATTACGCTAGGCGTATCTTTGATCATAGGGGCTTGCCTT 3726  
|||  
609 .....Leu 609  
3727 AACGACACCATACCTCAACTTGCTTACGTCCCAACCGGTAGAGGGAGAA 3776  
:;:;:|  
610 SerSerThr..... 612  
3777 TGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAG 3826  
612 ..... 612  
3827 GTGCTGGTCCATCTGGGACTCAGCATGAACGGTTCCCGGTGACGATAAC 3876  
612 ..... 612  
3877 GATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAATCTGGG 3926  
|||||  
613 .....SerValAspThrCysGluAspMetAspValThrGluAspMetG1 627  
3927 CGACGTGGGGATTAGCACCAACGGGGAAGTTACTGTGTGAAGAGAGACCC 3976  
|||  
627 uAspIle.....ValLeuMetAlaAspLysSerH 637  
3977 ATTCACTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTGCCCCG 4026  
|||||  
637 isSer.....TyrMetSerProGluMetAlaArg 646  
4027 TCTGCAAGCTCTGCGGAAAGACAAGGTGATCCGTGCGGTTCGTGACCGC 4076  
|||  
647 TrpAla..... 648  
4077 TGACGCTATGGCTTTTGTGAAGTGTGAAAAGGTGTCGACGATGTCT 4126  
:;:|  
649 .....AspValLysTyrGly..... 653  
4127 TTCACCAACAGTCTAGTGGGAAACGGCTCGTGAGTTGAGGTGGACGGC 4176  
:;:;:  
654 .....AsnAsn 655  
4177 AAAGGTTCTCCAGAAAGCTGTCGCTGAGGCGCCGACACAAAG 4226  
|||||  
656 LysGlyAlaLeuValGluTyrLysValGlyThrSerMetThr..... 669  
4227 GGAAGAGTGCAGATGTTGTAACACAGCACAAACCGCGGTCAACGAAGCG 4276  
669 ..... 669  
4277 ACAGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATT 4326  
:;:  
670 .....Leu 670  
4327 CCAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAGTGAACAAGA 4376  
|||  
671 Pro...AlaThrTrpAlaGlu.....LysValLysAlaVa 681  
4377 AGTACATTGGCGACTGTTTTCGGGGCCACGCCAATCGTCGATGAGAAAC 4426  
:;:|  
681 IleuProLeuSerGlyIleCysValArgLysPro..... 692

4427 CCGCCCAAGTGTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAG 4476  
692 ..... 692  
4477 GCGTCTCGCTCATGTGGCTGAGAAAAACAGGTACAACTCGACGAGCCCAA 4526  
|||  
693 .....GlnPheSerLysPro... 697  
4527 ACAGAGGAGTTTGACGATCAATGAAGCAAGCGCGTAAACAGCTTTGCA 4576  
697 ..... 697  
4577 TGTTTAGAACGCTGTTCTCGCGGTGTCAGCTGGATGTGTACAACGAAGCG 4626  
|||||  
698 .....LeuAspGluGluAspAsp... 703  
4627 ACTATCGCCACCAAGTGTCTCAACGCA...TTTACCTTTGTCGATAACTT 4673  
|||||  
704 .....LeuArgLeuSerAsnMetAsnPhePheLysValSerAspLe 717  
4674 GAAAGGAGGAGTGCCTCTTTTCTCAAAGCTGGGTGAGGGGTATACCT 4723  
|||  
717 uLys..... 718  
4724 ATAATGGTGGTAGCCATGTTTCATCAGGGTGGCTCGTGCCTTAGAGGAT 4773  
|||||  
719 .....LeuLysLys 721  
4774 ATCTTAAGGCAATTAAGTACCAAGCGTCTCGACCACTGTTTAGTGCA 4823  
:;:|  
722 ThrIleThrProValValThrGlyThrIleArgGluArgGlnMetLys 738  
4824 GAAGTACAGATGGGTGGAGCGGTACCATTCACGCTGATGACGAGGAGT 4873  
:;:|  
738 sAsnTyrIle..... 741  
4874 GCTATCAGATCAGATAACCCCTATCTTCAGGTCATCTCGTGGGAAGGCA 4923  
741 ..... 741  
4924 AACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGTGCATGTCATAAACGT 4973  
741 ..... 741  
4974 AGCTTCGGGTGACTATTTTCTTATGCCTTCGGTTCCTCAAGGACGCACT 5023  
|||||  
742 .....AspTyrLeuSerAlaSerLeuGlySerThrLeuGlyAsnL 755  
5024 TGCAATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGCTTCAGG 5073  
|||  
755 euGluArgIle..... 758  
5074 GCAACTCGGCGCTCTTGGTGTAGCAGGATGTTGCAGTTAGCCGCGG 5123  
758 ..... 758  
5124 CGTCTCGGATGAGAAGTCAACAGGTGTTCCAAACACGACCAACACAGAGCC 5173  
758 ..... 758  
5174 AAGGTCTACCAAGACAATACACCAAAATTCGGGGGCAAGGCTCTATCT 5223  
:;:  
759 .....ValArgSer 761  
5224 GAGGGAAGTGGTAGGAGTCAAGGGGAGGTGCACATCTCGATATGG... 5271  
:;:  
762 AspTrpAsnGlyThrGlu...GluSerMetGlnThrPheGlyLeuTyrAs 777  
5272 .TCGGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATC 5320  
|||||  
777 pCysGlu.....LysCysLysTrpLeu..... 784  
5321 CAGTGTAGTCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTT 5370

```
785      ||| ||| : : : : : |||||
      .....LeuLeuProAlaGluLysLysHisAlaTrpAlaValVal 797
5371 AAACCGGACCTCTGAAGATGCGCTGAGTACTTGAAGTACTGCG 5420
      : : : : : : : : : : : : : : : : : : : : : :
798 LeuAlaSerAspAspThrArgIlelle...PheLeuSerTyrAspG1 813
5421 TATAGGCATT .....GAGAGACATACAGGCGGTGCTTATGG 5458
      ||| : : : : : : : : : : : : : : : : : : : :
813 uSerGlySerProIleIleAspLysLysAsnTrpLysArgPheAlaValC 830
5459 CTAGAAATATTCGCTCACTACCGCGGAGGTGTTCTGAAAGTACCTAAT 5508
      : : : : : ||| : : : : : : : : : : : : : :
830 ysSerGluThrLysValTyrSerValIleArgSerLeuGluValLeuAsn 846
5509 CAAGTTTATGAATCACTACCGGGCTTCCAGTTCACAGTTCAGGCGGACAGA 5558
      : : : : : ||||| ||| : : : : : : : : : : : :
847 LysGluAlaIleValAspProGlyValHisIle..... 857
5559 TCTCATTTTTCATTCAACACAGAGCGGCTTCGCTGTGAGACCTACCGT 5608
      : : : : : : : : : : : : : : : : : : : :
858 .....ThrLeuValAspGlyVal..... 863
5609 ACGTACTCATAGCTGAAAGAGTATCTTTACCAAGGCAAGATGTCGAC 5658
863 ..... 863
5659 GCGGTGGTAGCTTTGGGCGACAATCTGTTGCTATCGGACGATATCTGGT 5708
863 ..... 863
5709 TTTCACGATGCCATTAAATTGATAGTGCACCTGAAAGTGCATCGATCG 5758
863 ..... 863
5759 GCATGGTGGCGAATCGTTTAAGTCTTCGAATAATAAGTGTATATGCT 5808
863 ..... 863
5809 CCCCAGGTGGGTAAGACGACGAGGTTAGTG ..... 5841
864 ...ProGlyCysGlyLysThrAlaGluIleIleAlaArgValAsnTrpLys 879
5842 .GACCAATTCTGTTAAGTCACCAATAGCACAGCCACCATACGGCTAATG 5890
      : : : : : ||| : : : : : : : : : : : : : :
879 sThrAspLeuValLeuThrProGlyArgGluAlaA..... 891
5891 TGGCAAGTTCGAGACATAAATATGCGCGTCAAGACAGAGATCCGAAT 5940
      : : : : : : : : : : : : : : : : : : : :
892 .....AlaMetIleArgArgAlaCysAla 900
5941 TTGGAAGGTCTCAACAGGTGCTACACAGTTAACTCCAGGTTGTAACCTT 5990
      : : : : : ||||| : : : : : ||| : : : : : : : :
901 LeuHisLysSerProValAlaIleAsnAspAsnValArgThrPheAspSe 917
5991 TATGCTC.....AGGGAATCTATAAA.....AGGCTTTTGGTGGATG 6028
      ||| : : : : : : : : : : ||| ||| |||||
917 rPheValMetAsnArgLysIlePheLysPheAspAlaValTyrValAspG 934
6029 AGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGCTTCGCAACC 6078
      ||| : : : : : : : : : : : : : : : : : : : :
934 LuGlyLeuMetValHisThrGlyLeuLeuAsnPheAlaLeuLysIleSer 950
6079 GGCGGTGGGAAGGCTCTTTTGGAGACATAAATCAGATACCATTTCAT 6128
      ||| : : : : : : : : : : ||||| ||| : : : : :
951 GlyCysLysLysAlaPheValPheLysAlaLysGlnIlePropheIl 967
6129 AAACAGGGAAGGTGTTAGCATGATGCTGCTGTTTTTGTCCAAAGA 6178
      ||||| ||| : : : : : : : : : : : : : : :
967 eAsnArg.....ValMetAsnPheAsp.....TyrProLysG 978
6179 AG.....GAAAGCTGTATACACTTCTAAATCGTAC 6210
      : : : : : : : : : : : : : : : : : : : :
```

```
978 luLeuArgThrLeuIleValAspAsnValGluArgTyrValThrHis 994
6211 AGGTGTCGTTAGATGTTTGTCTACTTGTCTCAATGACCGTAAGGG 6260
      ||||| ||||| : : : : : : : : : : : : : :
995 ArgCysProArgAspValThrSerPheLeuAsnThrIle..... 1007
6261 AACGGAAGGTGTACCTGAAAGGTGTTAGCGTAAGGACAAACCCAG 6310
      ||| : : : : : : : : : : : : : : : : : :
1008 .....TyrLysAlaAlaValAlaThrThrSerProV 1018
6311 TAGTAAGATCGCTGTCCAA.....AGG 6333
      ||||| : : : : : : : : : : : : : : :
1018 alValHisSerValLysAlaIleLysValSerGlyAlaGlyIleLeuArg 1034
6334 CCAATTGGAACCACTGATGACGTAGCTAGTAATAAACGCTGAGCTGACT 6383
      ||| : : : : : : : : : : ||| : : : : :
1035 Pro.....GluLeuThrLysIleLysGlyLysIle...Ii 1045
6384 GTGCATGACCCAGTTGGAGAGTCGGATATGAAGAGTCGTTGAAGGAA 6433
      : : : : : ||||| : : : : : : : : : : : :
1045 eThrPheThrGlnSerAspLysGlnSerLeuIleLysSer..... 1058
6434 AAGGAAAGAAACACCACTGATGACAGTCATGACGACGACGAAACACA 6483
      : : : ||| ||||| ||||| ||||| : : : : :
1059 .....GlyTyrAsnAspValAsnThrValHisGluIleGlnGlyLthr 1073
6484 TTCAGTGATGTGTATTGTTAGGAGCAAGAACGCGATGACTCCCTATT 6533
      ||||| : : : : : : : : : : : : : : :
1074 PheGluGluThrAlaValAlaGalaThrProThrProIleGlyLeuIi 1090
6534 CACTAAA...CAACCGCATATACTTTGTTGTTGTCGAGACACACACGCT 6580
      : : : : : : : : : : : : : : : : : : : :
1090 eAlaArgAspSerProHisValLeuValAlaLeuThrArgHisThrLysA 1107
6581 CACTGTTTATGCCCTCTGAGCTCAAGTTGGACGATAAGTCGCGCACA 6630
      : : : : ||||| : : : : : : : : : : : :
1107 laMetValTyrThrVal.....ValPheAspAlaValThrSer 1120
6631 TATATTAGCGGCGCTCACCTCAATCAGTATCCGACGCTTTCCTTCACAC 6680
      ||||| : : : : : : : : : : : : : : :
1121 IleIleAlaAspVal.....GluLysValAspGlnSerIleLeuThrMe 1135
6681 GTTCGCC 6687
      |||||
1135 tPheAla 1137
seq_name: pir2:JQ1896
seq_documentation_block:
hypothetical 191.1K protein - raspberry bushy dwarf virus
C:Species: raspberry bushy dwarf virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-2000
C:Accession: JQ1896
R:Ziegler, A.; Natsuaki, T.; Mayo, M.A.; Jolly, C.A.; Murant, A.F.
J. Gen. Virol. 73, 3213-3218, 1992
A:Title: The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus.
A:Reference number: JQ1896; MUID:93107856
A:Accession: JQ1896
A:Molecule type: mRNA
A:Residues: 1-1694 <Zi>
A:Cross-references: GB:S51557; NID:g262149; PIDN:AAB24599.1; PID:g262150
C:Genetics:
A:Map position: segment RNA-1
C:Superfamily: cucumber mosaic virus RNA 1 protein
C:Keywords: RNA replication
alignment_scores:
Quality: 291.00 Length: 477
Ratio: 1.282 Gaps: 15
Percent Similarity: 47.589 Percent Identity: 25.577
alignment_block:
```

US-09-301-906-3 x JQ1896

Align seq 1/1 to: JQ1896 from: 1 to: 1694

5350 CCAATGACATTTTGAAGTGGTTAAAACCCGGGACCTCTCTGAAGATGCGGTCGT 5399  
 725 ProGluThrProGluValThrMetThrGlyAsnProAsnAlaValIleAl 741  
 5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGACAGACACATACAGGCGGT 5449  
 741 aGluAlaIleSerIyrCysArgAlaGluValAspArgIleGlyLys... 756  
 5450 TGCATTATGGCTAGAAATATTCGCTCACTACCGCCGAAGGTGTCTGAAA 5499  
 757 .....LysCysGluArgIleLeu... 762  
 5500 GTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGCTC 5549  
 763 .....HisAlaIyrGlnAlaThrGlyAsnCysGlyGlyIyrLeuAs 776  
 5550 GGGCAGACATCTCATTTTTCATTCAACACAGACGGCTTGGCTGTGAGAG 5599  
 776 nAspThrAspAsnVal ..... 781  
 5600 ACCTACCGTAGCTCATCTACGCTGAAAGAAGTATCTTT ..... 5637  
 782 .....GlyValPheAspLysMetThr 788  
 5638 .....ACCAAGGCGAAAGATCTGACGCGGTGTACTTTGGCGGA 5678  
 789 SerTrpValGlnLysProLysGluPheAspHisGluPheGlyTrpAspG1 805  
 5679 CAATCTGTTCGTA .....TGCG 5695  
 805 ySerSerPheIleLysLeuSerTrpPheGlyLysIleProAspPheValG 822  
 5696 ACGATATACTGGTTTCCACGATGCCATTATTTGATTAGGTGCACTGAAA 5745  
 822 IyArgTrpLeuValValThrAspGlyThrArgValThrThrAsnLeuLys 838  
 5746 GTCGCTCGATCGCGATGGTGGCGGAATCGTTTAAGTCTTCGAATAATAA 5795  
 839 PheSerArgGlnIyrAlaThrIleProAlaThrValThrProThrIleLys 855  
 5796 GTGCTATTAATGCTCCCGCAGTGGGGTGAAGACGACGACGTTAGTGGAGC 5845  
 855 sLeuValAspGlyValThrGlyCysGlyLysThrGluIleValArgA 872  
 5846 AATTCGTTAAGTCACCCCAATAGCACGCCACCATTCGCGTATGTTGGGA 5895  
 872 rGyTrpArgProGlyIleLeuIleLeuSerValCysLysAlaAsnVal... 887  
 5896 AGTCTTGAGGACATAAATATGCGGTGAAGACGAGATCCGAAATTTGGA 5945  
 888 .....AspGluIleArgArgLysLeuAlaAlaValAspSerLysPhe... 901  
 5946 AGGTCTCAACAGTCTACCACAGTTAATCCAGCGGTGGTAACTTTATCG 5995  
 902 .....IleArgThrValAspSerTrpLeuLeuSerProSerV 914  
 5996 TCAGGGGAATGTATAAAGGGTTTTTGGTGGATAGGTGCACATCATCATCAT 6045  
 914 alThrGlySerCysAspLuleuPheIleAspGluIyrGlyLeuSerHis 930  
 6046 CAAGCGTTTACTACAACCTAGGCGTCTTCCCAACCGCGCGTCCGAAGCGCT 6095  
 931 ProGlyIleLeuLeuLeuAlaIleHisIleSerGlyIleArgLysValTh 947  
 6096 CTTTTTTTCGACACATAATTCAGATACCATTCATAAACAGGAGAGAGTGT 6145  
 947 rLeuPheGlyAspSerGluGlnIleProPheCysAsnArg ..... 960  
 6146 TTAGGATGGATGTGCTGCTGTTTTTCTTCCAAAGAGGAAAGCGTTT... 6189

961	.....LeuAlaaspPheProLeuLysTyTrAsnSerValGluasp	973
6190	.....GTATACACITCTAAATCGGTACAGGTCTCC	6218
974	ValGlyLeuAsnPheAspArgGluLeaArgSerThrThrTyArgCysPr	990
6219	GTATAGATGTTGCTACTGTTGTTGCTCTCAATGACCGTAAGGGGAACGGAAA	6268
990	oGlnAspIleThrLeuSerLeuGlnLysMet.....TyrL	1002
6269	AGTGTTCACCTGAAAGAGTCGTTAGCGGTAGGACAAACAGTAGTAGA	6318
1002	ysThrLysProIleLysThrValSer.....ThrValGluSer	1014
6319	TGCGTGTCCAAAGGCCAATTTGGAACCACTGATCACGTACTGACGTGAATATAA	6368
1015	SerIleThrIleLysProIleLysSerGluPheGluLeuProLeuProAs	1031
6369	CGCT.....GACGTGTACTTGTGTCATGACCCAGTGTGGAGAAGT	6406
1031	nAlaPheAspGlyProValLeuTyIleCysMetThrLysHisAspGluS	1048
6407	CGGATATGAAGAGTCTGTTGAAGGAAAGAAAGAAACACACAGTCATG	6456
1048	erLeuLeuLysLeuAArgTrpAlaLysGluAsnIleSerSerGluValArg	1064
6457	ACAGTGCATGAAGCAGCGGAAAAACATTACGTGATCATGTGTTATTGTTAG	6506
1065	ThrValHisAlaIaGlnGlyLeuSerTyrrLysAsnValValtyrPheAr	1081
6507	GACGAAGAAAGCGGTACTCCCTATTCACTAAACAA.....CCGCATA	6550
1081	gLeuThrArgThrAspAsnAspLeuTyThrLysArgLysLeuProtyrH	1098
6551	TACTTGTGTTGTTGTCGAGACACACAGCTCACTGGTTTATGCGGCTCTG	6600
1098	IsLeuValaIleSerArgHisThrAspLysIleValTyrcysThr...	1113
6601	AGCTCAAAAGTTGACGATTAAGTCGGCAGCATATATTAGCCAGCGGTCACC	6650
1114	.....ThrLysProGluaspSerSerAs	1121
6651	TCAAATCATGATCCGACGCTTTGTTCTACACG	6681
1121	pPheSerLeuSer...AlaLeuLysAsnThr	1130

seq\_name: p1rl:p13vbb

seq\_documentation\_block:

RNA la protein - broad bean mottle virus (strain Bawden)  
 N:Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-) methyltransferase  
 C:Species: broad bean mottle virus  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 05-May-2000  
 C:Accession: A41699  
 R:Dzianott, A.M.; Bujarski, J.J.  
 Virology 185, 553-562, 1991  
 A:Title: The nucleotide sequence and genome organization of the RNA-1 segment in two  
 A:Reference number: A41699; MUID:92074218

alignment_scores:		
Quality:	290.00	Length: 775
Ratio:	0.773	Gaps: 33



6109 ATAAATCAGATACCACTTCATTAACACGAGGAAGCGTGTTTAGGATGGAT... 6156

788 ThrGluGlnIleSerPheLysSerArgAspAlaThrPheArgMetLysTy 804

6157 -TGTCGCTGTTTTGTTCCTCAAGAAAGAAACGGTGTATACACTTCTAAAT 6205

804 rCysAsnIleGluTyrAspLysArgAspIleVal.....SerLysT 818

6206 CCGTACAGGTGCGGTATAGATCTTGGCTACTGTGTGTCCTCAATGACGGTA 6255

818 hrPheArgCysProGlnAspVal.....ValSerAlaValLysIle 831

6256 AGGGGAACCGGAAAAGTGT.....TACCCTGAAAAGGT 6287

832 Leu...LysArgLysCysAlaAsnArgSerSerLysTyrAsnGlyTrpVa 847

6288 CCGTTAGCGGTGAAGCAAAACCCAGTAGTAAAGATCGCTCTCCAAAGGCCAA 6337

847 lSerSerLys.....ValGluLysSerLeuSerLysSerArgI 861

6338 TTGGAACCACTGATGACCTAGCTAGCTGAAATAAACCGCTGACGTGTACTTGC 6387

861 leValSerIleAsnGlnValSerMetGluLysHisLysPheTyrLeuThr 877

6388 ATGACCCAGTTGGAGAAGCTCGGATTAAGAGAGTCTGTTGAAGGGAAGAAAG 6437

878 MetThrGluAlaAspLysAlaLeuCysSerArgAlaLysAspValGI 894

6438 A.....AAGAAACACCATGATGCATGACAGTCGATGCAAGCAC 6472

894 yLeuAspLysThrTrpValGluSerAsnMetGluThrValHisGluAlaG 911

6473 AGGGAACCAATTCAGTGTGCTGTGTTGTTAGGACGAAGAAGCCGAT 6522

911 lnGlyLysAlaValAspHisValValLeuValArgLeuLysSerThrLys 927

6523 GACTCCCTTAATTCATAACCAACCGCATATATCTGTGTGGTTTGTGCGACGA 6572

928 CysAspLeuPheLysSerGluGluTyrCysLeuValAlaLeuThrArgH1 944

6573 CACAGCTCACTGCTGTTTCCCGCTCTGACCTCAAGCTTGACGACGATAAGG 6622

944 sLysArgThrPheGluTyrLeuTyrAsnGlyAspLeuGlyAspLeuI 961

6623 TCGGCACATATATTAGC 6639

961 leSerPheTyrValThr 966

seq\_name: p1r2:JO2183

seq\_documentation\_block:  
 hypothetical 216.5K protein - apple chlorotic leaf spot virus (isolate P-205)  
 N:Alternate names: ORF 1 protein  
 C:Species: Apple chlorotic leaf spot virus  
 C:date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 20-Jun-2000  
 C:Accession: JQ2183  
 R:Sato, K.; Yoshikawa, N.; Takahashi, T.

J. Gen. Virol. 74, 1927-1931, 1993  
 A: Title: Complete nucleotide sequence of the genome of an apple chlorot  
 A: Reference number: JQ2183; MUID: 93389448  
 A: Accession: JQ2183  
 A: Molecule type: mRNA  
 A: Residues: 1-1885 <S>  
 A: Cross-references: GB: D14996; NID: g434059; PIDN: BAA03641.1; PID: g453240  
 C: Superfamily: eggplant mosaic virus RNA-directed RNA polymerase  
 C: Keywords: ATP

```
alignment_scores:
  Quality: 281.00
  Ratio: 0.466
  Percent Similarity: 43.041
  Percent Identity: 20.271
  Gaps: 74
  Length: 1401
```

```

alignment_block:
US-09-301-906-3 x J02183      ..
Align seg 1/1 to: J02183 from: 1 to: 1885

2938 TCTAAGGGTTTCGTACGTCGTGGCGTGCACGAAGCCCTTTTCTGCCAT 2987
|||||.....:|||||.....:|||||.....:|||||.....
235 SerSerIleLeuValArgAsnLysAlaThr.....GlyVa 247

2988 TAGCTCGTTCTTTCCACAGGTTCACTATTACGACCGCGGTTTA.... 3033
:.....:|||||.....:|||||.....
247 IGluIleArgTyrGlnVal...SerLeuValTyrSerLeuGlySerHisH 263

3034 .....ACTGAAGATGAAGAAGCTTGATGCTCTGGTG 3063
|||.....:|||||.....
263 IsLeuPheHisIleTyrProThrGluAsp.....LeuMet 274

3064 CCGCAGAGAAGTCGTATAAACTCACCGGTGGGCATACGTGGAGACGTCGCG 3113
::: ||| .....:|||||.....
275 LysGluGluValArgArgPheGlyProTyrAspLeuPheAspValGlySe 291

3114 CGTAGCTGTGACAAGTCGTAGCTGGAAACGAAAGAATTTTGAGATGAAG 3163
::: Phe::: ||| .....:|||||.....
291 rLeuPheValLysProVal.....ArgV 299

3164 TTTCCTTAATGACITTCACCACCTTTTCGTATTCCGGAATAGGTGCTTATC 3213
|| .....:|||||.....
299 alProIleGlnAspPheProLeuSerValPheLysLys.....Ile 312

3214 GGGATTATCGTCGGCGTCTTGGTGCGGCCCAATTGCATGGAAGTATAG 3263
|||||.....:|||||.....
313 PheIleTyrLeuSerSerLeuLysLysProAspVal..... 324

3264 GCGCGGAATTGCGGCTAACGCTAGAGGTACGCGGCAGCTAGTTACGAAA 3313
324 .....:|||||..... 324

3314 CTCTAAGCTCGTTAAGTTTCACAAAGCGCGGTGGTTTACGCGGTTTAAAC 3363
:::::::::::|||||.....
325 .....GlnSerAlaValAlaLysLeuArgGlnLeuSer 335

3364 TCTAGCACAGTATCCGTTGGTCTTAGTCGTGCGAAGAGGGTTTTCGTC 3413
::::: ||| |::::: ||| |::::: ||| |::::: ||| |::::: |||
336 AspAlaAspIleSerIleGluSerValPheMetValGlnGluPheAlaSe 352

3414 GCGGTGACCGTCACTAGGGCG.....:|||||..... 3435
| ||| ||| ::: :::
352 rArgValGluLysAsnGlyValGlySerTrpSercySerPheTrpGluC 369

3436 .....ACCGTAGCTAAAGCTCAAGTCCCCTTA 3462
::: ||| :||||| |||
369 ysMetLysAspTrpPheAspLysLeuProTyrArgGluVal...Leu 384

3463 GCGTTGCTATCCTTTCTACCTCATACGCC.....ATTTCGCGCTGCAG 3506
::: ||| :||||| |||
385 GluLysIleGlyLeuAlaAsnAspPheThrArgArgLeuMetLysIleLy 401

3507 TATGTTAGGCATTTGGGCACATGCTCTCCACGGCCTTAATGTTTTTCT 3556
: |||::: ||| |::::: ||| |::::: ||| |::::: |||
401 sProLeuAlaPheAspIleHisThrAspArgProLeuThrValArgM 418

3557 TTGGTTTAGGGACATTGCTTGGGCGAGGGCTAGCGCAATCTTGGGAAG 3606
::: ||| :||||| |||
418 etValIleAspGlnIleTrpGlyGluArgGlnSerSer..... 430

3607 TTTGGAGGCTCTCCAATAATTGGTGC...GCTGTTCCGAGGTTGTTTG 3653
|||||.....
431 .....CysaspAspValProAsnIleIlePh 439

3654 GCGAGGGAAGACT.....GTCAAGCTCATTTGTTACTGCCTATTACGC 3694
: |||::: |||
439 eTyrGlyArgLysGluTrpLeuGluHisGlyLeuLeuProLysValLysL 456

```



[illegible]

1140 euMetLeuLysMethisMetGluGly. .... 1148

6106 GACATAAATCAGATACCACTTTCATAAACACGGGAGAAGCTCTTTAGGATGGA 6155

1148 ..... 1148

6156 TTGTGCTGTTTTTTGTTCCAAAGAAAGCGTTGTATAC. .... 6195

1149 .....1leLeuLysLysSerThrileValCysilleGlyAspp 1161

6196 ..ACTTCTAAATCGTACAGGTGTCGGTTAGATGTTTGCTACTTGTTGTCTCC 6243

1161 roLeuGlnAlaGlyTyrrPheCysProLysAspAsnTyrrLeuSerArg 1177

6244 TCAATGACCGTAAGGGGA. ....ACGAAAGGTGTTA 6275

1178 GluGluGluLeuLysArgLeuPheLysAspGlyIleValAsnTyrrLysTrpTy 1194

6276 CCCTGAAAGGTGCGTTAGCGGTAGGACAAACCGACTAGTAGATCGCTGT 6325

1194 rSerTyrrArgile. ....AsnLysPheilleAlaLysLysLeuA 1207

6326 CCAAAAGGCCAAATTCGACACCACTGATGAGCGACTCTGCAATTAACCGCTGAC 6375

1207 la.....IleGluThrMetAsnAspPheilleGlyIleAspGluGln 1220

6376 GTGTACTTGTGCATGACCCAGTGGAGAAGTCGGATATCAAGAGGTCG.. 6423

1221 .....AsnLeuIleTyrrLysAspMetProSerAlaHL 1231

6424 .....TTGAAGGAAAGAAAGAAACACCA. .... 6450

1231 sHisPheMetCyluAsnLysGlyAsnHisIleGluValIleLeuValProS 1248

6451 .....GTGATGACAGTGT 6462

1248 erMetIleGluLysGluLeuTyrrSerAsnTyrrGlyAsnValMetThrPhe 1264

6463 CATCAAGCACAGGAAAAACATTCAGTGATGGGTATCTCTTTAGGACGAA 6512

1265 GlyCluserGlnGlyLeuThrPheAsnCysGlyIleValIleValLeuSerGI 1281

6513 GAAAGCCGTGACTCCCTATTCTCACTAAACCGCATATACTTGTGGTT 6562

1281 uGluAlaLys. ....LeuCysSerAspAlaHisIleMetValAlaI 1295

6563 TGTCGACACACACCGCTCACTGTTTATGCGGCTCTGAGCTCAAG 6609

1295 leThrArgPheArgGlyGlyPheCysPhe..AlaLeuGlySerLys 1309

**This Page Blank (uspto)**



Sequence	Strd Orig	zScore	EScore	Len	Documentation
SwissProt_39_RRPO_SHVX +	414.00	445.71	1.9e-18	1718	! Q04575 shallot virus x (shv)
SwissProt_39_RRPO_PMW +	404.00	435.47	7.9e-18	1547	! P20591 papaya mosaic potexvirus
SwissProt_39_R194K_TRVSY +	338.50	359.55	1.2e-13	1707	! P05080 tobacco rattle virus
SwissProt_39_RRPO_SHMV +	337.00	358.36	1.5e-13	1629	! P89202 sunn-hemp mosaic virus
SwissProt_39_R1V25_AWVLE +	329.50	353.98	3.8e-13	1126	! P03589 alfalfa mosaic virus
SwissProt_39_RRPO_TOM51 +	313.00	331.00	5.0e-12	1616	! Q9YKd6 tomato mosaic virus
SwissProt_39_RRPO_TOML +	309.00	336.48	8.9e-12	1616	! P03587 tomato mosaic virus
SwissProt_39_RRPO_TOMK2 +	303.00	319.62	2.5e-11	1616	! P89676 tomato mosaic virus
SwissProt_39_RRPO_TOMK1 +	302.00	318.48	2.5e-11	1616	! Q9g1t8 tomato mosaic virus
SwissProt_39_RRPO_CMVJS +	298.00	313.69	4.5e-11	1648	! P19523 cucumber green mottle
SwissProt_39_RRPO_PMWJV +	297.00	312.80	5.2e-11	1612	! P89657 pepper mild mottle virus
SwissProt_39_RRPO_PMWVS +	297.00	312.80	5.2e-11	1612	! P29098 pepper mild mottle virus
SwissProt_39_V1A_BMVV +	290.00	310.61	1.1e-10	966	! Q00020 broad bean mottle virus
SwissProt_39_RRPO_THVRA +	288.00	302.49	1.9e-10	1616	! Q98745 tobacco mosaic virus
SwissProt_39_RRPO_THVB +	285.00	299.06	3.0e-10	1616	! Q9Yq08 tobacco mosaic virus
SwissProt_39_RRPO_ACLS3 +	281.00	292.75	5.8e-10	1885	! P54891 apple chlorotic leaf
SwissProt_39_RRPO_ORSVS +	279.00	292.24	7.2e-10	1612	! Q84133 odontoglossum ringsp
SwissProt_39_RRPO_ORSVV +	277.00	289.95	9.7e-10	1612	! P89659 odontoglossum ringsp
SwissProt_39_V1A_PSVJ +	275.50	293.60	9.7e-10	1005	! P28726 peanut stunt virus (sc
SwissProt_39_V1A_BMV +	275.00	293.53	1.0e-09	961	! P03588 bromo mosaic virus (bm
SwissProt_39_RRPO_THVM0 +	275.00	287.64	1.3e-09	1616	! Q93058 tobacco mosaic virus
SwissProt_39_V1A_CMVN +	274.00	292.02	1.2e-09	993	! P17769 cucumber mosaic virus
SwissProt_39_V1A_BSMV +	273.50	289.89	1.4e-09	1139	! P17595 barley stripe mosaic
SwissProt_39_V1A_CMVNT +	273.00	290.88	1.4e-09	993	! Q40976 cucumber mosaic virus
SwissProt_39_V1A_CMVO +	273.00	290.88	1.4e-09	993	! P20122 cucumber mosaic virus
SwissProt_39_RRPO_THGVW +	272.50	284.84	1.9e-09	1609	! P18339 tobacco mild green m
SwissProt_39_FIG2_YEAST +	270.50	282.55	2.5e-09	1609	! P25653 saccharomyces cerevis
SwissProt_39_RRPO_THOB +	270.50	282.50	2.5e-09	1616	! P90211 tobamovirus ob. rna-r
SwissProt_39_V1A_CMV11 +	270.00	287.45	2.2e-09	993	! Q83270 cucumber mosaic virus
SwissProt_39_V1A_CMV1X +	269.00	286.31	2.5e-09	993	! Q66121 cucumber mosaic virus
SwissProt_39_RRPO_THV +	267.00	278.51	4.2e-09	1616	! P03586 tobacco mosaic virus
SwissProt_39_RRPO_THVKR +	264.00	275.08	6.5e-09	1616	! P30738 tobacco mosaic virus
SwissProt_39_V1A_CMVQ +	263.00	279.48	6.0e-09	991	! P06011 cucumber mosaic virus
SwissProt_39_RRPO_ACLS3P +	261.00	269.91	1.1e-08	1884	! P27738 apple chlorotic leaf
SwissProt_39_V1A_CMVY +	253.50	268.61	2.4e-08	993	! Q83264 cucumber mosaic virus
SwissProt_39_V1A_CMW +	252.00	267.30	3.0e-08	958	! P27752 cowpea chlorotic mottle
SwissProt_39_VJ3P_YEAST +	241.50	253.13	1.5e-07	1161	! P47179 saccharomyces cerevis
SwissProt_39_AMTH_YEAST +	240.00	249.56	2.0e-07	1367	! P08640 saccharomyces cerevis
SwissProt_39_RRPO_TVCV +	238.00	245.49	2.9e-07	1601	! Q88920 turnip vein-clearing
SwissProt_39_V1A_TAV +	237.00	249.76	2.7e-07	993	! P28931 tomato aspermy virus (ta



```

965 MetThrPheGlnLysAlaCysGluArgGlu.....CysLysSe 977
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5829 GAGCAGCTTAGTCGACGAATTCCTTAAAGTCACCAATAGCACAGCCACCA 5878
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
977 rValThrIlePheAspTyrGlyLysLeuPro.....AlaGlyPheV 992
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5879 TTACGGCTTAATGTGGCAAGTCTCGAGACATAAATATGCGGTG..... 5922
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
992 aAspAlaTyrLeuAlaIleLysValAsnValGluLeuAlaIleLeuThr 1008
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5923 .....AAGAAGAGAGATCCGAATTTGGA 5945
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1009 GlyAspGlnArgGlnSerThrHisHisGlnGluArgGluSerGlnIleSe 1025
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5946 AGGTCCTCAACAGTGTACACAGATTAACTCCAGGGTGTAAACTTTATCG 5995
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1025 rSerLeuGlnSerAsnIleAlaGlnPheSerLysTyrAlaAspTyrL 1042
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5996 TCAGGGGAATGTAT..... 6009
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1042 euAsnAlaThrHisArgGlnProArgArgLeuAlaAsnProIleLysVal 1058
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6010 .....AAAAGGTTTGTGTGGA 6026
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1059 HisAlaGluArgGlnLeuGlyGlyAlaValLeuLysAlaAsnIleValPr 1075
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6027 TGAGTGCACATGATG.....CATCAAGGCTTACTA.. 6057
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1075 oAspLeuAlaMetValLeuValProAlaPheArgSerGlnSerLeuLeuT 1092
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6058 ..CAACTAGCGCTTCGCAACC.....GGCGCGTCGGAAGCCCTTT 6099
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1092 hrAspLeuGlyArgHisAlaMetThrTyrAlaGlyCysGlnGlyLeuThr 1108
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6100 TTGGAGACATAAATCAGATA.....CCATTCTATAACAG 6134
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1109 LeuAsnHisLeuThrIleLeuAspLysAspThrProLeuCysSerAs 1125
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6135 GGAGAAGGTGTTTAGCATGATTGCTGCTTTTGTTCCTCAAGAGAA 6184
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1125 pGluValLeuTyr.....ThrAlaPheSerArgAlaSerGluS 1138
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6185 GCGTTGTATACACT.....TCTAATCTGTACAGGTCTCCG 6219
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1138 erIleThrPheValAsnThrHisSerAspAsnProAlaPheLeuAlaLys 1154
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6220 TTAGATGTTGTTGCTACTTGTGCTCCTCAATGACC.....GTAAGGGGAAC 6263
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1155 LeuAspAlaThrProTyrLeuLysThrLeuIleSerTrpValArgGluAs 1171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6264 GGAATAAGTCTTACCCTGAAAGGTCCTTAGCGGTAAAGGACAAACCAAGTAG 6313
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1171 pGluGlu.....AlaGlyAlaAspCysProAlaT 1181
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6314 TAAGATCGCTGTCCAAAGG.....CCAATTGGACCACTGATGACGTAGCT 6360
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1181 hrGluProLeuValLysAspValProThrLysThrHisIleProValAla 1197
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6361 GAAATAACCGCTGACGTGACTTGTGTCATGACCCAGTGTGAGAAGTCGGA 6410
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1198 AsnAspLysValGlnLeuGluGlyLysIleGluAlaMetGluAspLysAs 1214
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6411 TATGAAGAGGTGCTTGAAGGGAAGAAAGAAAGAACCAACCACTGATGACAG 6460
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1214 pThrArgGluLeuTrpSerGlyGluGluLysThrAsnLeuMetGlnThrG 1231
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6461 TGCATGAACACAGGGAAGAAACATTTCAGTGTGTTGTTTATTGTTAGACG 6510
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1231 InAspProVal.....ValGlnLeuPheProHis 1240
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6511 AGAAAGCC...GATGACTCCCTATTTCATCAACACCCGCATATACTTGT 6557
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

1241 GlnGlnAlaLysAspGluAlaLeuPhe..... 1249
5558 TGCTTTGTCGAGACACACACGCTCACTGGTTTATGCGGCTCTGAGCTCAA 6607
1250 ..... 1250
6608 AGTTGACGAGTAAGTCCGACATATATTAGCGACGCGTCCACCTCAATCA 6657
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1250 ysIle.....ThrIleGlyGluArgIleArgMetAlaThrProGluGln 1264
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
6658 GTATCCGACGCTTGTCTTCCACACG 6681
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1265 AsnAlaLysGlnLeuArgHisThr 1272
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: SwissProt_39:RRPO_PMW
seq_documentation_block:
ID RRPO_PMW STANDARD; PRT; 1547 AA.
AC P20951;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICATION PROTEIN (176 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
DE DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
OS Papaya mosaic potexvirus (PMV);
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=89381685; PubMed=2778435;
RX Sit T.L., Abouhalidar M.G., Holy S.;
RT "Nucleotide sequence of papaya mosaic virus RNA.";
RL J. Gen. Virol. 70:2325-2331(1989).
CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13957; BAA03050.1; -
DR PIR; JQ0096; JQ0096.
DR INTERPRO; IPR000606; -
DR PFAM; PF01443; Viral_helicase1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transference.
FT NP_BIND 822 829 ATP (POTENTIAL).
SQ SEQUENCE 1547 AA; 176320 MW; 6CA4282C6A082622 CRC64;
alignment_scores:
Quality: 404.00 Length: 798
Ratio: 1.101 Gaps: 35
Percent Similarity: 45.990 Percent Identity: 23.559
alignment_block:
US-09-301-906-3 x RRPO_PMW ..
Align seg 1/1 to: RRPO_PMW from: 1 to: 1547
4157 GTGAGTTGTGAGTGACGCGCAAGGTTGCTCCAGAAAGCGTCGTCGCT 4206
      ||| ||| :::::  :::::  :::::  :::::  :::::  :::::  ||
405 ValThrLeuThrTyrProThrLysThrPheAspAsnArgGlyTrpValva 421
4207 GAGGCGCCGACACAGAAAGGGAAGAGCTGCAGATGGTAACACAGCACA 4256
      ||:::  :::::  :::::  :::::  |||
421 LlysLeuGluAlaArgArgGlyTyrGluTrpPheAlaAspGluMethIsL 438
4257 ACCCGGGTCAACGAAG..... 4273
      ||||| :::::  :::::  :::::  :::::  :::::  :::::  :::::

```



```

5881 ACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGCGGTGGAAGAG 5930
910 IleIleuThrGlyAspCys.....LysGI 918
5931 AGATCCGAATTTGGGAAGTCTCAACAGTGTCTACACAGTTAACTCCAGG 5980
918 nSerValHisHisGluSerAsnGluAsnAlaThrThrSerSer..... 932
5981 TGTAACTAATTTATCGTCAGGGAATGTATAAA.....AGGGTTTGGTG 6024
933 ..IleGluProLeuValGlyGluAlaSerGluLeuCysArgTyrTyrIle 948
6025 GATGAGGTGCACATGATGATCAAGCTTACTA....CAACTAGCGCTCTT 6071
949 AsnAlaThrHisArgAsnLysLysAspLeuAlaAsnLysLeuGlyVal 965
6072 CGCA.....ACCGCGCGCTCGAA..... 6090
965 rSerGluLysThrGlyLeuThrGluValThrHisGlyThrThrProIleP 982
6091 ..GGCTCTTTTGGAGACATAAATCAGATACCATTCATTAACAGGAG 6138
982 roGlyLeuHisMet.....LeuValProSerLeuTyrLysLys 994
6139 AAGGTGTTTAGG.....ATGATTTGTC 6161
995 GlnAlaPheSerGluMetGlyHisLysValSerThrTyrAlaGlyCysGI 1011
6162 TGTGTTTGTTCACAAAGAAAGCGTCTATACACTTCTTAATCGTACA 6211
1011 nGlyIleThrAlaProLysIleGlnIleLeuLeuThrGluGluThrSerL 1028
6212 GGTGTCGGTTAGATGTTCTACTTCTCTCTCAATACCGGTA 6255
1028 euCysSerArgGluValLeuTyrThrAlaLeuSerArgAlaVal 1042

seq_name: SwissProt_39:194K_TRVSY

seq_documentation_block:
ID 194K_TRVSY STANDARD; PRT; 1707 AA.
AC P05080;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last annotation update)
DE POTENTIAL 194 KDA PROTEIN (PUTATIVE REPLICASE) [CONTAINS: 134 KDA
DE PROTEIN].
OS Tobacco rattle virus (strain SYM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88034943; PubMed=3668507;
RA Hamilton W.D.O., Boccara M., Robinson D.J., Baulcombe D.C.;
RT "The complete nucleotide sequence of tobacco rattle virus RNA-1.";
RL J. Gen. Virol. 68:2563-2575(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX Boccara M., Hamilton W.D.O., Baulcombe D.C.;
RT "The organisation and interval homologues of genes at the 3' end of
RT tobacco rattle virus RNA1.";
RL EMBO J. 5:223-229(1986).
CC -1- SIMILARITY: THIS PROTEIN IS HOMOLOG TO PUTATIVE REPLICASE OF
CC TOBACCO MOSAIC VIRUS (TMV) AND TRICORN VIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06172; CAA29537.1;
CC EMBL; D00155; BAA00110.1;

```

```

DR PIR; S01865; S01865.
DR INTERPRO; IPR000606;
DR INTERPRO; IPR001788;
DR INTERPRO; IPR002588;
DR PFAM; PF00978; RNA_dep_RNAPol2; 1.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
FT CHAIN 1 1187 134 KDA PROTEIN.
FT CHAIN 1 1707 194 KDA PROTEIN.
SQ SEQUENCE 1707 AA; 194328 MW; F4DABDB6C7E51F04 CRC64;

alignment_scores:
Quality: 338.50 Length: 1858
Ratio: 0.486 Gaps: 72
Percent Similarity: 37.513 Percent Identity: 17.976

alignment_block:
US-09-301-906-3 x 194K_TRVSY
Align seg 1/1 to: 194K_TRVSY from: 1 to: 1707
1333 GACGCAATGATTCGCGAAAAACCTCA.....CATAG 1364
50 AspAspMetIleArgGluArgLysThrLysAspProValIleValHisGI 66
1365 GTCCGATGTGTTTCTGAAGCGCGGAAAGGAGAGAGCTGAGGAACCTCT 1414
66 u.....ValLeuSerGlnLysGluGlnAsnLysLeuMetGluIleT 80
1415 TTCCAGAGCTTTTCGATACAGTTCTCCGACTCGGTGAGGAGTAGTCACCCA 1464
80 yrProGluPheAsnIleValPheLysAspLysAsnMetValHisGly 96
1465 TTCCGTAATGCCATCGCGAGCTCTTTCAATGGAATCTTTTCCAGGAGCTG 1514
97 PheAlaAlaGluArgLysLeuGlnAlaLeuLeuLeuLeuAspArgVa 113
1515 TGCTAATGTGTGC...TTCTTCGATATTGGGGGAGCTTCACGTATCATG 1561
113 lProAlaLeuGlnGluValAspAspIleGlyGlyGlnTrpSerPheTrpV 130
1562 TCNAAGCTGGCCATGTGAAGTGTATGTCATGTCATCAATCAGTCCTAGAGTT 1611
130 alThrArgGlyGluLysArgIleHisSerCysCysProAsnLeuAspIle 146
1612 AAAGATGTGAAGCGGAGATCAATGAGATCCTCTTTTCCACAGCTGG 1661
147 ArgAspAspGlnArgGluIleSerArgGlnIlePheLeuThrAlaIleGI 163
1662 GCGAGATTCCTACGCTCCGACTGACCTTCTTAAGTAA..... 1698
163 yAspGlnAlaArgSerGlyLysArgGlnMetSerGluAsnGluLeuTrpM 180
1699 .....GCGGCTTCAAAAGTCTGTGTCTTACTGT 1725
180 etTyrAspGlnPheArgLysAsnIleAlaAlaProAsnAlaValArgCys 196
1726 AGTCGAGAATCCGAGAACTCCGATTCTAGACCGCATCGCGGTTT..... 1770
197 AsnAsnThrTyrGlnGlyCysThrCysArg.....GlyPheSerAs 210
1771 .....ATGGTGGATGTGTACG 1786
210 pGlyLysLysLysGlyValaGlnTyrAlaIleAlaLeuHisSerIleTyrA 227
1787 ATATATCCCGCAGCAGGTAGCAGAGCTATGGATAAGAGGCTGCGCTG 1836
227 spPheLysLeuLysAspLeuMetAlaThrMetValGluLysLysThrLys 243
1837 GTTTTCCACATACCTCTTATGTTTC...CCGCTGAGTGTGTGTCAGGTAA 1883
244 ValValHisAlaAlaMetLeuPheAlaProGluSerMetLeuValAspGI 260

```

```
1884 CGTGAAGCTTTACTTGAAGAACTCGATACGTTGGTGAAGAGGGAAGGTG 1933
:|||||
260 uGly.....ProLeuProSerValAspGlyT 269
1934 ATTACCTGGCTACAAATGTTGGTCAGTGGTGAGATGTATGAAACATTCC 1983
|||||
269 yTyrMetLysLysAsn.....GlyLysIleTyr..... 278
1984 TTCCTCAACGTAAGCGGTTTTTC.....ACCTTTTCTTATGTACG 2024
|||||
279 .....PheGlyPheLLeuLysAspProSerPheSerTyrIleHI 291
2025 CACTTCGTCCGGGAACGTTGTTAAGCTAGATGATGAGGATACCGTTGTG 2074
:|||||
291 sasPrpGlu.....GluTyrLysLysTyrLeuLeu. 301
2075 GTTACCATCATCTCACTATGTAGGTGTCAGAACTCACCTGGGAAGTACG 2124
|||||
302 .....GlyLysPro 304
2125 GTTACGTATAGTCTGTCCTCCGTCGTTCTGGGCAAAATCGCTGGTGT 2174
|||||
305 ValSerTyrGln.....GlyAsnValPheTyrPh 314
2175 CATACCT...GTTGTAGCTGGTTCTAGTGTGTCTCTTAAAGACAATAGTCC 2221
|||||
314 eGluProTrpGlnValArgGlyAspThrMetLeuPheSer..... 327
2222 TCGATTCCGGACTTTGTCACAGGATCTATTCTCAGCGCTCAACACTATA 2271
|||||
328 .....IleTyrArgIleAla..... 332
2272 GGGACATTCGAGATAGAACGTTTCAGTATGCCGTTGGGCGGTCAAGTTC 2321
|||||
333 .....GlyValProArg... 336
2322 GCAAAAGACCATGTGTATTACAGGAGTGGCGTTGTCACACAGCAAGGTTG 2371
336 ..... 336
2372 ATATTTCTCTGATGATATGTTGGGGTTTAGTGTGCGGTATGGCTCAG 2421
336 ..... 336
2422 CCGATTAAAGGATAGGGCGAAGATTCGCTCCTATAACTTTATAAAGC 2471
|||||
337 .....ArgSerLeuSerSerGlnGluTyrTyrArg... 346
2472 CAGTGAGGGAGTCTCGCGGGGCTTCAAGCTCTTCTTTTCAGACCGTAG 2521
346 ..... 346
2522 GCGATTGTTTTTCGAACGCACTCCCGTCTATGCTAAGGCAATGGTGCAC 2571
346 ..... 346
2572 GATAACTTCAACGTTTTCGACACGCTTATGTCTATGCCCCAGCGTTTCAT 2621
|||||
347 .....ArgIleTyrIL 350
2622 CCGTAAAGTACCTGGTCTGTTGTTGTATTCACATTTGCACCTTCTGGAGCTT 2671
|||||
350 eSerArgTrpGluAsnMetValValProIle..... 361
2672 CAGACAGTTGGAGCTCAGGGGTGCTTTGATATTTGGAAGAGACCTTC 2721
|||||
362 .....PheAspLeuValGluSerThr... 368
2722 GGTAGGAAACTGAAGAATAGTCGCTTGGCGCTCTCTCTAGGGCTATCGT 2771
|||||
369 .....ArgGluLeuVa 372
```

```
2772 GGAAGATTCAATTAAGGTCTATGAAGCAATGAAGACAGAAGATGGAAC 2821
|||||
372 lLysLysAspLeuPheValGluLysGlnPheMet..... 383
2822 CCCTCCCAATTACTCAAGATCTGTATATGCGTTTCATTAATGGGG..... 2865
|||||
384 .....AspLysCysLeuAspTyrIleAlaArgLeuSer 394
2866 .....AACGTTTCTAACGTCCTACTGTACGAGGCGAGGCTTCTTCT 2903
|||||
395 AspGlnGlnLeuThrIleSerAsnVal..... 403
2904 TGGCGGTTTCGAAAGCGACCGTGTTCGAGTGTTCCTAAGGTTTGGTAG 2953
|||||
404 .....LysSerTyrLeuSerSerAsnTrpValLeuPheI 416
2954 CTCGTGGGCTCCGACGAGGCGCTTTCTGGCATTACGTCGTTCTTTCC 3003
|||||
416 leAsnGlyAlaAlaValLys..... 422
3004 ACAGGTTTCACTATTCTACGACCGCGTTTAACTGAAGATGAAAGGCTTCA 3053
|||||
423 .....AsnLysGlnSerValAspSerArgAspLeuGI 433
3054 TGCTCTGTGGTGGCACA.....GAGAATGCTATAAACTCACCGGTGG 3094
|||||
433 nLeuLeuAlaGlnThrLeuLeuValLysGluGlnValAlaArgProVa.. 449
3095 GCATCTCGAGACGTCGGCGTAGCTGTGACGCAAGGTGCTAGCTGGAACG 3144
|||||
450 .....MetArgGluLeuArgGluAlaIleLeuThrGluThr 461
3145 AAAGAATTTTGGAGTGAAGTTTCCCTTAATGACTTCACCACCTTCGTAT 3194
|||||
462 LysProIle.....ThrSerLeuThrAsp..... 469
3195 CGGGAATAAGGTGCTTATCGGGATATTCTGGTGGCTTTTGGGTGGCGCC 3244
|||||
470 .....ValLeuGlyLeuIleSerArgLysLeu..... 478
3245 CAATTGCATGGAAGTATAGGCGCGGAATTCGGGCTAACGCTAGAAGGTAC 3294
|||||
479 .....TrpLys.....GlnPhe 482
3295 GCGGCGCAGTAGTTACGAACTCTAAGCTCGTTAAGTTTCACAGCCGCGG 3344
|||||
483 Ala.....AsnLysIleAlaValGI 489
3345 TGGTTTACGCGGTTTAACTCTAGCACAGTATCCGGTGGATCTTTAGTCG 3394
|||||
489 yGlyPheValGlyMetVal.....GlyThrLeuIle. 499
3395 TCGGAAGAGGTTTTCGTGGGCGGTGACCGTACACTAGGGCGACCTAGCT 3444
|||||
500 .....GlyPheTyrProLysLysValLeuThr..... 508
3445 AAACGTCAGTCCCTTAGCGTTGCTATCGTTTCTACCTCATACGCCAT 3494
508 ..... 508
3495 TTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACT 3544
|||||
509 .....TrpAlaLysAspThrProAsnGly. 516
3545 TAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGCGTAGCGG 3594
516 ..... 516
3595 AATACTTGAAGCTTTGGAGGCTTCTCCAATAATTTGGTGGCTGTTCCCGA 3644
|||||
517 .....ProGI 518
3645 GGTTGTTGGGAGGGAAGAGTGTGAGCTCATTTGTTACTGCTATTAGGC 3694
```

```

1: 518 uLeuCysTyrGluAsnSerHisLysThrLysValIleValPheLeuSerV 535
1: 3695 TA.....GGGGTATCTTTGATCATAAAGGGCTTGGTTAAC 3729
1: 535 alValTyrAlaIleGlyGlyIleThrLeuMetArgArgAspIleArgAsp 551
1: 3730 GACACCATCCTCAAGTTCGTTACGTCGCCACCGGTAGAG...GGAGGAA 3776
1: 552 GlyLeuValLysLysLeuCysAspMetPheAspIleLysArgGlyAlaH1 568
1: 3777 TGTGTACGAT.....GAGACCGTTCAGGTATTACCGGGACTTTGACTATG 3820
1: 568 sValLeuAspValGluAsnProCysArgTyrTyr..... 579
1: 3821 ACAGAGTGTGTCTCCATCTGGGACTCAGCATGAAGGGTTCCCGGTGAC 3870
1: 580 .....Glu 580
1: 3871 GATACGATGGATCCACTTCTAGTCTCAAGCTATCATCTTGTGCACAAA 3920
1: 581 IleAsnAspPheSerLeuTyrSerAla..... 591
1: 3921 TGTGCGCGACGTGGGATTAGCACCAACGGGGAAGTTACTGTGAAGAAG 3970
1: 592 .....SerGluSerGlyGluThr..... 597
1: 3971 AGACCCATTCACTCGAAGCGTCAATACACTTATCTCGAGGAAGAGTT 4020
1: 598 .....Val 598
1: 4021 GCCCGCTCTCACCTGTGCGGAAACACAAAGGTGATCCGTCGGTCTCG 4070
1: 599 LeuProAspLeuSerGluValLysAlaLysSerAsp..... 610
1: 4071 TACCGCTGACGCTATGCTTTCTTCAAGTGTCAAAAAGGTGTCGACG 4120
1: 611 .....LysLeuLeuGlnGlnLysLysGluIleAlaAspG 622
1: 4121 ATGCTCTTCCACCAAGTCTAGTGGGAAACGGCTCGTGAGTTGAGTG 4170
1: 622 LuPheLeuSerAlaLysPheSerAsnTyrSerGlySerValArgThr 638
1: 4171 CACGGCAAAGGTTGCTCCAGCAAAACGGTCGTCGGTCAGCGCCGACACA 4220
1: 639 Ser.....ProProSerValValGlySerArgSerGI 650
1: 4221 AGAAGGGGAAGCTGCAGATGTTAAACAGCACACAAACCGCGTCAACG 4270
1: 650 yLeuGlyLeuLeuGluAspSerAsnVal..... 660
1: 4271 AAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCCGCCACAGCT 4320
1: 660 ..... 660
1: 4321 GATATTCCAAAGTCAACCGTCCGAGGTACATGCTCAGAAAGATG... 4368
1: 661 .....LeuThrGlnAlaArgValGlyValSerArgLysValAs 673
1: 4369 .AAACAAGAGTACCATTGCGGACTCTTTCGGCGCCACGCCCAATCGTCG 4417
1: 673 pAspGluGluIleMetGluGlnPheLeuSerGly..... 684
1: 4418 ATGAGAAACCCGCCCAAGTGTGTACGACTCGTGTGTGAAGATAATTGAC 4467
1: 685 .....LeuIleAsp 687
1: 4468 AAGGCGAAGCGCTCCTCATGTGGCTGAGAAAAACAGGTACAACTCGA 4517
1: 688 ThrGluAlaGluIleAspGluValSerAlaPheSerAlaGluCysGI 704
1: 4518 GCAGCCCAACACAGAGGAGTTGACGATCAATGAAGCGCAAGCGGTAAAC 4567
1: 704 uArg.....GlyGluThrSerGlyThrLysV 713

```

```

4568 AGCTTTGCATGTTTAGAACGTGTTCTCGCGGTGTGCAGCTGGATGTGTAC 4617
713 alLeuCysLysProLeuThr..... 719
4618 AACGAACGCACTATCGCCACCAAGGTTCTCAACGCCAATTACCTTTGTCGA 4667
719 ..... 719
4668 TAACTTGAAGAGGAGGAGTGGGTCTTTTCTCAAAGCTGGGTGAGGGGT 4717
719 ..... 719
4718 ATACCTATAATGGTGTAGCCATGTTTTCATCAGGGTGGCCTCGTGCCTA 4767
720 .....ProProGlyPhe 723
4768 GAGGATATCTTAACGGCAATTAAAGTACCCCAAGCGTCTTTCACCACTGTTT 4817
724 GluAsnValLeuProAlaValLys..... 733
4818 AGTGCAGAAGTACAAGATGGGTGGAGCGGTACCATTCCACGCTGATGACG 4867
733 uValSerLysGlyLys..... 738
4868 AGAGTGTCTATCCATCAGATAACCCCTATCTTGACGGTCAATCTCGTGGGG 4917
739 .....ThrValLysArgVal... 743
4918 AAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGAAGTCAATGTCAT 4967
743 ..... 743
4968 AAACGTAGCTTCGGTGACTATTCTTTTCTTTCCTTGGGTTT..... 5010
744 .....AspTyrPheGlnValMetGlyGlyGluArgLeuP 755
5011 .CAAAAGGCGCACTTGCATTCACTAACTCCATCGACGAAGGCGGCATC 5058
755 roLysArgProValValSerGlyAspAsnSerValAspAlaArgArgGlu 771
5059 AGTTTCAGCTTCAGGCAACTCGCGCGTCTTTTGGTGTAGGACAGGATGTT 5108
772 PheLeuTyrTyrLeuAspAlaGluArgValAlaGlnAsnAspGluIleMe 788
5109 GCAGTTAGCGCGCGGTGTCGGATGAGAGTCAACAGGTGTTCCAAACC 5158
788 tSerLeu.....T 791
5159 AGCAACCAACAGAGCCAAAGTGTCTACCAGAAACAATCACACCAAAATCGGG 5208
791 yrArgAspTyrSerArgGlyValIleArgThr.....Gly 802
5209 GGCAAGGCTCTATCTCAGGGAAGTGGT.....AGGGAAGTCAAGGG 5249
803 GlyGlnAsnTyrProHisGlyLeuGlyValTrpAspValGluMetLysAs 819
5250 GAGGTCGACATCTGATATGTTGGCGAACAAGATTACGTTAGGAAGTGTG 5299
819 n.....TrpCysIleArgProValValThrGluHis 830
5300 AGTGGCTCAGGGCTGATATCCAGTGTGCTTGAACCTGACTACACC 5349
830 latyrVal.....PheGlnProAsp..... 836
5350 CCAATGACATTTGAAGTGTAAACCGGGACCTCTGAAGATGCCGTGCT 5399
837 .....LysArgMetas 840
5400 GGAGTACTTCAAGTATCTGCTATAGGCATT...GAGGACATACAGGG 5446
840 paspTrpSerGlyTyrLeuGluValAlaValTrpGluArg..... 853

```

```
5447 CGTTCTTATGCTAGAAATATTGCCGTCACCTACGCCGAAGGTGTTCTG 5496
853 .....
5497 AAAGTACCTAATCAAGTTTATGAATCACTACGGGCTTTCAGCTTTACAA 5546
853 .....
5547 GTCGGGCACAGATCTCATTTTTCATTCAACACAGACGGCTTTCGCTGTA 5596
854 .....GlyMetLeuValA 858
5597 GAGACCTACCGTACCTACTCATAGTGAAGGATATCTTTACCAAGGC 5646
858 snAspPhe..... 860
5647 AAAGATGTCGACGGGTGGTAGCTTTGGCGGACAAATCTGTCGTATGCGA 5696
861 .....AlaValGluArgMetSerAspTyrValIleValCysAs 873
5697 CGATATACTGGTTTCCACGATGCCATTAAATTGATAGTGCACCTGAAG 5746
873 pGlnThrTyrIeuCysAsnAsnArgLeu...IleLeuAspAsnLeuSerA 889
5747 TCGCTCGATGCGCGCATGCTGGCGGAATCGTTTAACTCCTTCCAATATAA 5796
889 laLeuAspLeuGlyProValAsnCysSerPhe.....Glu 900
5797 TGCATATAATGTCCTCCCGAGTGGCGGTAAAGACGACGCTTAGTGGACGA 5846
901 leuValAspGlyValProGlyCysGlyLysSerThrMetIleValAsn... 916
5847 ATTCGTTAAGTCACCCATAGACACAGCCACCATACGCTTAATGTGGGA. 5895
917 .....SerAlaAsnProCysValAspValValLeuSerThrGlyA 930
5896 .....AGTCTCAGGACATAAAT 5913
930 rgAlaAlaThrAspAspLeuIleGluArgPheAlaSerLysGlyPhePro 946
5914 ATGGCGGTGACAGACAGAGATCCGAATTTGGAAGCTCTCAACAGTGCTAC 5963
947 CysLysLeuLysArgArg.....Vally 954
5964 CACAGTTAACTCCAGGGTGTAAACTTTATCGTCAGGGAATGTATAAA 6013
954 sThrValAspSerPheLeuMetHisCysValAspGlySerLeuThrGlyA 971
6014 GGGTTTTC...GTGCGATGAGTGCACATGATCATCAAGGCTTACTACAA 6060
971 spValLeuHisPheAspGluAlaLeuMetAlaHisAlaGlyMetValTyr 987
6061 CTAGCGCTTCTCGCAACCGCGCTCGGAAGCCCTCTTTTGGAGACAT 6110
988 PheCysAlaGlnIleAlaGlyAlaLysArgCysIleCysGlnGlyAspG1 1004
6111 AAATCAGATACCATTCATAAACAGGAGAGAGTGTTTAGGATGGATTGTG 6160
1004 nasCnGlnIleSerPheLysProArgValSerGlnValAspLeuArgPheS 1021
6161 CTGTTTTCCTCCAAAGGAAGGAGCGTGTATACACTTCTAAATCGTAC 6210
1021 erSerLeuValGlyLysPheAspIleValThrGluLysArgGluThrTyr 1037
6211 AGCTGCTCGTTAGATGTTTCTACTTGTTCCTCAATGACCGTAAGGG 6260
1038 ArgSerProAlaAspValAlaAlaValLeuAsn..... 1048
6261 AACGGAAAGTGTACCTGAAAGGTCGTTAGCGGTAAAGGACAAACCCAG 6310
1049 .....LysTyrTyrThrGlyAspValArgThr.....HisAsnAlaT 1061
```

```
6311 TAGTAAGATCGCTGTCCAAAAGGCCAAATTTGGAACCACTGATGACGTAGCT 6360
1061 hrAlaAsnSerMetThrValArgLysIleValSerLysGluGlnValSer 1077
6361 GNAATAAACCGTGCAGCTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
1078 LeuLysProGlyAlaGlnTyrIleThrPheLeuGlnSerGluLysG. 1094
6411 TATGAAGAGGTGCTTGAAGGAAAGAAAGAAACACCAAGTCATGACAG 6460
1094 uLeuValAsnLeuLeuAlaLeuArgLysValAlaAlaLysValSerThrV 1111
6461 TGCATGAACACAGAGGAAACATTCAGTGTGTGTATTTAGGAGC 6510
1111 alHisGluSerGlnGlyGluThrPheLysAspValValLeuValArgThr 1127
6511 AGAAAGCCGATGACTCCCTATTCTACTAACAACCGCATATATCTTCTGG 6560
1128 LysProThrAspSerIleAlaArgGlyArgGluTyrLeuIleValAl 1144
6561 TTGTGCGAGACACACAGCTCACTGTTTCTGCGCTCTGAGCTCAAGT 6610
1144 alLeuSerArgHisThrGlnSerLeuValTyrGluThrValLys..... 1158
6611 TGGACGATAAGTCGGCACATATATTAGCGACGCTCACCTCAATCAGTA 6660
1159 ..GluAspValSerLysGluIleArgGluSerAla.....AlaLeu 1172
6661 TCCGACGCTTTCCTTCACAGCTTC 6684
1173 ThrLysAlaAlaLeuAlaArgPhe 1180
seq_name: SwissProt_39:RRPO_SHMV
seq_documentation_block:
AC RRPO_SHMV STANDARD; PRT; 1629 AA.
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (128 KDA PROTEIN)].
OS Sunn-hemp mosaic virus (SHMV) (TMV strain cowpea).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE OF 1-1536 FROM N.A.
RX MEDLINE=97093421; Pubmed=8938983;
RA Silver S., Quan S., Deom C.M.;
RT "Completion of the nucleotide sequence of sunn-hemp mosaic virus: a
RT tobamovirus pathogenic to legumes.";
RL Virus Genes 13:83-85(1996).
RN [2]
RP SEQUENCE OF 1538-1629 FROM N.A.
RX MEDLINE=83064542; Pubmed=6292867;
RA Meshi T., Ohno T., Okada Y.;
RT "Nucleotide sequence of the 30K protein cistron of cowpea strain of
RT tobacco mosaic virus.";
RL Nucleic Acids Res. 10:6111-6117(1982).
CC -I- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -I- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -I- MISCELLANEOUS: REATHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR LYS-1130 AND GLN-1132.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



DR	ENBL; U47034; AAB38492.1; -;
DR	ENBL; J02413; AAA46585.1; -;
DR	INTERPRO; IPR00606; -;
DR	INTERPRO; IPR001788; -;
DR	INTERPRO; IPR002588; -;
DR	PRAM; PF00978; RNA_dep_RNApol2; 1.
DR	PRAM; PF01443; Viral_helicaseI; 1.
DR	PRAM; PF01660; Vmethyltransf; 1.
KW	Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT	CHAIN 1 1629 RNA-DIRECTED RNA POLYMERASE.
FT	CHAIN 1 1130 METHYLTRANSFERASE/RNA HELICASE.
NP_BIND	849 856 ATP (POTENTIAL).
SQ	SEQUENCE 1629 AA; 186513 MW; C92B8C9AF5FE8A5 CRC64;

alignment_scores:		
Quality:	337.00	Length: 1898
Ratio:	0.504	Gaps: 73
Percent Similarity:	35.248	Percent Identity: 18.335

alignment block:

US-09-301-906-3 x RRPO\_SHMV ..

Align seq 1/1 to: RRPO SHMV from: 1 to: 1629

1279 TCGGCA<sup>nt</sup>TGCTTCCCA<sup>nt</sup>TTGGACACGAAATTGTCGATATGTTTCTCA 1328  
 |||::|||::|||:: |||::|||:: |||:: |||::  
 4 SerThrLeuIleAsnLysAlaGlnThrAsnSerCysGlyAspValGlyVa 20

1329 AAGGACGCAATGATTCGGAA..... 1350

20 1valAspLeuLeuLysArgLysvalTyrAspAspThrvalLysThrmetG 37

1351  
AAACCTCACATAGGTGGGATGTGTTCTGACGCC 1386

37 | nc | v | eu | as | na | r | q | a | l | v | e | t | r | a | r | q | e | u | a | s | n | c | n | c | u | f | o | u | c | u | p | r | o | 52

[illegible][illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

.....

[illegible][illegible][illegible]

104 ASPRICEGUYNSURNEFROATIANISMELMELLYSEYATYSEYIVA 120

1384 TCATGTAATGCAATCCAGATCCATAGACGTTAAAGATGTGAAGCGGAGAATCA T633

120 I H I S C Y S A S N P R O A I A L E U A S P A I A A R G A S P L E U A I A R G A S N G L U A 137

1634 ATGAGATCCTCTTCTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGT 1683

137 snTyrArgIleSerIle.....Glu 143

1684 GACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTGTAGT... 1728

144 AsnTyrLeuSerArgPheGluAspLysSerGlyAspTyrCysGlnTrpG1 160

1729 .....CGAGAATCGCAGAACT 1744

160 nArgLysLysProLysValSerLysProLeuProArg<sup>Tyr</sup>GlnLysAlaC 177

1745 GCGATTCTAGAGCCGAT..... 1761

177 yspheAspArgTyrAsnGluAspProGluHisvalThrCysSerGluThr 193

1762 .....GC 1766  
194 PheGluLysCysArgIleSerProProAlaGluArgAspAspIleTyrAl 210  
1764 GGGTTTTATGGTGGATGTTACGATATATCCCGCAGCAGGTAGCAGAG 1813  
210 aThrSerLeuHisSerLeuTyrAspIleProTyrGlnAsnLeuGlyProA 227  
1814 CTATGGATAAGAAGGGTCGCTGTTTTCGCATACAGCTCTTATGTTCCCC 1863  
227 laLeuAlaArgLysArgIleLysValLeuHisAlaAlaPheHisPheSer 243  
1864 GTGAGTGTGTTGACCGTAACGGTGAAGTTTACTTTGGAAGAACTCGATAC 1913  
244 GluAspLeuLeuGlyAlaSerGluGlyLeuLeuThrGlnIleGlyG 260  
1914 GTTGGTGAAGGAAGGTGATTACCTGGCTACAAAT...GTTGGTCAGT 1960  
260 yThrPheGlnArgAsnGlyAspValLeuThrPheSerPheLeuAspGluS 277  
1961 GTGTGGAGATGTTATGAACATCTCTCTTAACAGTAAAGGGGTTTTTCAC 2010  
277 erSerLeuIleTyrThrHisSerPheArgAsnValPheGluTyrValThr 293  
2011 TTTTCTTATGTACCACTTCCTCGGGGAACGTGTTTAAAGCTAGAGTATGA 2060  
294 ArgThrPhePheValAlaCysAsnArgTyrAlaTyrMetLysGluPheAr 310  
2061 GGGATACCGT.....TGTTGGTTACCATCATCTC...ACTA 2092  
310 gSerArgArgValAspThrValPheCysSerPheIleArgIleAspThrT 327  
2093 TGTGTAGGGCTCAGAAGTCACTCGAACGTAGGTTACGTATAGGTCGTTG 2142  
327 yrCys.....LeuTyrArgSerVal 333  
2143 GTCCCGTCGTTGCTGGGCAATCGCTGGTGTACACCTGTTGTAGCTGG 2192  
333 ..... 333  
2193 TTCTAGTGTCTCTTTAAGACAATAGTCTCGATTGGCACTTTGTCGACA 2242  
334 .....PheLys.....AspCysAsp..... 338  
2243 GGATCTATTCTACGGCTCAACACTATAGGACATTTCGAGAATAGAACG 2292  
338 ..... 338  
2293 TTTTGTATATCCGTTGGGGCGGTCAAGTTCGCAAAAGACCCATGTCTATAC 2342  
339 ..GluHisValPheAlaAlaMet..... 345  
2343 AGGAGTCGGTGTCTCACACAGCAGGTGATATTCTCTGATGATATGT 2392  
346 .....AspAspAlaT 349  
2393 GGGGTTTA...GTTGTCGCTCTTATGGCTCAGCGCATTAAGGATAGGGCG 2439  
349 rpgLluPheLysLysArgValMetLeuGluAlaSerArgProIlePhe 365  
2440 AAGAGTATTCGCTCTATACTTTTAAAGCCAGTGAAGGGAGTCTCGC 2489  
366 AsnAspValAlaGlnPheAsnValTyr..... 374  
2490 CGGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCAACG 2539  
375 .....PheProAsnA 378  
2540 CAGTCTCCGCTATGCTAAGGCATGGTCAGGATAACTTCAACGTTTTG 2589  
378 la..... 378  
2590 GAGACGCTTATGCTATGCCAGACGGTTTCATCCGTAAGTACCTTGGGTC 2639

```

379 LysAspLysValCysLeuProIlePheAlaValLysSerValSerGlyAl 395
    ::: ::: ||| ::::::::::||| ||::
2640 TGTGTGTACCATTTGCACCTTCTGGAGCTTCACAGAGTTGGAGCTCA 2689
    ||:::|||
395 aProValThr 403
    |||
2690 GGGGTGCCTTGATATTCGAAGAGAGACTTCGGTAGAAACTGAAGAT 2739
    :::::::::::
404 .....ValGluLysAspPheTyrTrpThrAlaLeuAsnHis 415
    :::::::::::
2740 ACTCCCTTCGGCCTCTCTAGGCTATCGTGGAGATTCAATTAAAGT 2789
    :::::::::::
415 ..... 415
2790 CATGAAGCAATGAAGACAGAAAGATGGAAACCCCTGCCAATTACTGAAG 2839
    ::::::::::
416 .....IleLeuThrTyrProAspGlyLys 423
    ::::::::::
2840 APTCTGTATATCGTTTCATAATGGGAACGTTTCTAACGTCCTCACTGTACG 2889
    ::::::::::
423 ..... 423
2890 AGGGCAGTCTCTTGGCGGTTCGAAGAGGACCGTGTTCGAGTGTTC 2939
    ::::::::::
423 ..... 423
2940 TAAGGTTTGGTAGCTCGTGGGCTGCGACGAAGCCCTTCTGGCAATTA 2989
    ||| ||| |||
424 .....AlaAspPheArgGlyValM 430
    ::::::::::
2990 CGTCGCTCTTTCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAA 3039
    ||| ||| |||
3040 GATGAAGCTTGTAGTCTCGTGGCGCACAGAGATGCTATAAACTACC 3089
    ::::::::::
436 .....IleArgSerArgValValIleAsn 443
    ::::::::::
3090 GGTGGGCATCTGGAGAGCTGCGCGTAGCTGAGCAAGGTCGTAGCTG 3139
    ||| ||| |||
444 ....GlyThrThrAlaSerGlnTrpGluValAspLys 455
    ::::::::::
3140 CAACCAAGAAATTTGGAGTGAAGTTTCTTAAATGACTTCACCCTTC 3189
    ::::::::::
456 .....Ser 456
3190 GTATTGCGGAATAGGTCTTATCGGGATATTCGTGGCGTCTTTGGTGC 3239
    ||| ||| |||
457 GlnLeuLysAspIleAlaLeuSerLeuLeuIleAlaLysLeuGluLys 473
    ::::::::::
3240 GCGCCCAATTCGATGGAGTATAGCGCGGATTCGGCTAACGCTAGNA 3289
    ||| ||| |||
473 sLeuLysIleSerValIleGluLysArg 482
    ::::::::::
3290 GGTACCGCGGAGTAGTTACCAACTCTAAGCTCGTTAAGTTTCAAGCC 3339
    :: :: :: ||| ||| |||
483 .....IleLysIleGluArgGlnGlyLeuValSerLeuLysLysLysPhe 497
    ::::::::::
3340 CCCGTTGTTTACGCGGTTTAACTCTAGCAGCATATCCGGTGGATCTTT 3389
    ||| ||| |||
498 LeuHisGlyLeu 512
    :: :: :: :: ::
3390 AGTCGTGGAGAGGTTTTCGTGGCGGTGACCGCTCACTAGGGGACCG 3439
    ||| ||| ||| ||| |||
512 pValValGluLysGlyTyrPValLysSerValAsp 525
    ::::::::::
3440 TAGCTTAAACGTCAGTCCCTTAGCGTTGCTATCGTTTCTACCTCATAC 3489
    || :: :: :: ||| ||| |||
525 aIleGlnValThrIleProAspLeuValLeuAsnPheArgAspHisPhe 541
    :: :: :: :: ::
3490 GCCATTTCGGCTGCAGATATTTAGGCATTTTGGGCACATGCTCTTCCACG 3539
    ||| ||| |||
```

```

542 Arg 545
3540 GCACCTAATGTTTTTCTTTAGGACATTGCTTGGGGCGAGGGCTA 3589
    ||| ||| |||
546 .....ArgThrS 548
3590 GCGCAATACACTTGGAGTTTGGAGGCTTCTCCAATAATTGGTGCCTGTT 3639
    ||| ||| |||
548 erAlaAsn 551
    ::::::::::
3640 CCCGAGGTTTGGGAGGAGAGTGTCACTCATCTTACTGCCTAT 3689
    ||| ||| |||
552 SerGluVal 562
    ::::::::::
3690 TAGCTAGGGGTATCTTTGATCATAGGGGCTTGCTTAACGACACATAC 3739
    |||
563 .....ThrAsnG 565
3740 CTCAACTTCTTACGTCCACCGTAGAGGGAGGAATGTGTACGATGAG 3789
    ||| ||| |||
565 lutyTyrAlaLysValSerAspLeuValAspArgAsn 578
    ::::::::::
3790 AGGCTTAGGTATTACGGGACTTT 582
    ||| ||| |||
579 ThrLeuAlaPhe 593
    ::::::::::
3828 TGCTGCTCATCTGGGACTCAGCATGAAGCGTTCCCGGTGACGATAACG 3877
    |||
593 sLeuGly 595
3878 APTGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCG 3927
    :: :: :: ||| ||| |||
596 .....ValAspIleAspThrValThrGluLeuIle 605
3928 GACGTGGGATTAGCACCAACGG 612
    ||| ||| |||
606 Asp 621
    ::::::::::
3963 TGAAGAAGAGACCCATTCACTCGAAGCGT 638
    ::::::::::
621 pAspLysGluGlnLeuProArgThrLeuAlaGlySerSerSerTyrLeu 638
    ::::::::::
4007 TCGAGGAAGAGGTGCTCCCGCTGTCAGCTGCTGGCGGAAGACAAAGTGAT 4056
    ::| ||| |||
638 euGluGlu 641
    ::::::::::
4057 CGTGGGTCTGTGATACCGCTACCGCTATGGCTTTTGTGAAGTGTGAA 4106
    ||| |||
642 ProSer 643
    ::::::::::
4107 AAAAGTGTGACGATGCTTTTCCACACAGCTAGTGGGGAACGCTC 4156
    ||| ||| |||
644 .....AspAspLeu 646
    ::::::::::
4157 GTGAGTTGAGTGGACGGCAAAAGGTTGCTCCCAAGAAC 4203
    ||| ||| ||| |||
647 .....ValCysLeuSerAspLysAlaIleValAsnArgSerThrIleLeu 661
    ::::::::::
4204 GGTGAGGCGCCGACACAGAAAGGGAGAGCTGCAGATGGTAACACAGC 4253
    ||| ||| |||
662 GlyGluLeuLysAsn 666
    ::::::::::
4254 ACAACCGCGTCAACGAGGACAGGGAGCGAGTACAGTCCAGTCTTTG 4303
    :: :: :: ||| ||| |||
667 AsnValValIlePheGluGlyThrLeu 682
    ::::::::::
4304 TGAGTCCGCCACAGCTGATATTCAAAAGGTCAAC 4347
    ||| ||| ||| |||
682 aIserAlaPro 696
    ::::::::::
4348 GTACATGTCTCAGAAAGAGTGAACAAAGATACCATTTGCCGACTGTTTC 4397
    ::| ||| |||
697 LeuHisAlaArg 700
    ::| ||| |||
```

4398 GGGCGCCACGCCAATCGTCGAT.....GAGAACCCGCCCCAA 4435  
|||||  
701 .....ProValSerAspPheLeuSerMetGlnLysProValAsnI 714  
4436 GTGTTACGACTCGTGGTGAAGATAATTGACAAAGGCAAGGCCGTCGCT 4485  
||| ||| |||  
714 leValTyrThrGlyGluValGlnIle..... 722  
4486 CATGTGGCTGAGAAAAACAGGTACAAAGTCGACGCCAACCAACAGAGGAG 4535  
722 ..... 722  
4536 TTGACGATCAATGAGGCAAGCCCGGTAAACAGCTTTGCATGTTAGAA 4585  
722 ..... 722  
4586 CGTGTTCCTCGGTGTCAGCTGGATGTGTACAACGAAGGACTATCGCC 4635  
|||  
723 ..CysGlnMetGlnAsnTyrLeuAsp...TyrLeuSerAlaSerLeuVal 737  
4636 ACCAGGTTCTCAACACCATTTACCTTTGTGCACTAACTTGAAGGGAGGAG 4685  
::: |||  
738 AlaCysIleSerAsnLeuLysLysTyrLeuGlnAspGlnTrpLeuAsnPr 754  
4686 TGGCGTCTTTTCTCAAGCTGGGTGAGGGGTATACCTATATATGTTGGTA 4735  
::: |||  
754 oGlyGluLysPheGlnLysIleGlyVal.....TrpA 765  
4736 GCCATGTTTCATCAGGCTGCGCTCGTCCCTAGAGGATATCTTAAACGGCA 4785  
765 spAsnLeuAsnAsnLysTrp..... 771  
4786 ATTAAGTACCAAGCGCTCTCGACCACTGTTTAGTCAGCAAGTACAAGAT 4835  
||| ||| |||  
772 IleValValPro.....GlnLysLysLysTy 780  
4836 GGGTGAGCGGTACCATTCACCGCTGATGACGAGGAGTGTATCCATCAG 4885  
::: |||  
780 rAlaTrpGlyLeuAlaIleAspValAspGlyAsnGln..... 792  
4886 ATAACCCCTATCTGACGGTCAATCTCGTGGGAAGGCAAACTTCTCGACT 4935  
792 ..... 792  
4936 AAGTCAGGAAGGGTGGTAAGGTCAATGTCATAAACGTAGCTTCGGGTGA 4985  
|||  
793 .....LysThrValIleLeuAsnTyr..... 799  
4986 CTATTTTCTATGCTTGGCTTGGTTCCTTCAAGGACGCACTTGCATTCAAGTAA 5035  
799 ..... 799  
5036 ACTCCATCGACGAAGGCGCATCAGTTTGACGTTTACGGGCAACTCGCGCG 5085  
799 ..... 799  
5086 GTCTTTGGTGTAGGCAAGGATGTTGACGTTAGCGGGCGGCTGTCGGATGA 5135  
|||  
800 .....AspGI 801  
5136 GAAGTCACCAAGGTGTTCCAAACCAAGCAACCAACAGCAGCCAAAGGTGCTACCA 5185  
|||  
801 uHIs.....GlyMetPro..... 805  
5186 GAACAATCACACCAAAATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGT 5235  
805 ..... 805  
5236 AGGGAAGTCAAGGGGAGGTTCGACATACTCCATATGTTGCGCAACAGATTAA 5285  
|||  
806 .....IleLeuGluLysSerty 811

5286 CGTTAGGAAGTGTGAGTGGCTCAGGCTGATATCCAGTGCATGGCTCTTG 5335  
|||||  
811 rValArgLeu..... 814  
5336 AACCTGACTACACCCCAATGACATTTGAAGTGTAAACCCGGGACCTCT 5385  
|||||  
815 .....ValValSerThrAspThrTyr 821  
5386 GAAGATGCCGTCGTCGAGTACTTGAAGTATCTCGCTATATAGGCATTGAGAG 5435  
::: |||  
822 LeuPheThrValValSerMetLeuGlyTyrLeuArg..... 833  
5436 GACATACAGGCGCTGCTTATGCTAGAAATATTTGCGTCACTACCCCGG 5485  
833 ..... 833  
5486 AAGGTGTTCTGTGAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTT 5535  
833 ..... 833  
5536 CACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGG 5585  
||| |||  
834 .....HisLeuAspGln..... 837  
5586 CTTGCGTGTGAGAGACCTACCGTACGTACTACTAGTGAAGAGGTATCT 5635  
837 ..... 837  
5636 TTACCAAGGCAAAAGTGTGACGCGGTGCTAGCTTTGGCGCACAACTCG 5685  
||| |||  
838 .....LysLysProThrAlaThrIleThrLeuValaspGlyVal 850  
5686 TTCGTATGCCACGATATCTGCTTTTCCACGATCCCATTAATTTGATAGG 5735  
850 ..... 850  
5736 TGCACGTGAAGTCGCTCGATGCGGCATGTTGGCGGAATCGTTTAAGTCCT 5785  
850 ..... 850  
5786 TCGAATATAAGTCTATATGCTCCGCCAGTGGCGGTGAAGACGACGACG 5835  
||||| |||  
851 .....ProGlyCysGlyLysThrGlnGlu 858  
5836 TTAGTCGACCAATTCGTTAAGTCACCCCAATAGCACAGCCACCATTAAGGC 5885  
::: |||  
859 IleLeuSerArgPhe.....AspAlaAsnSerAspLeuIleLeu..... 871  
5886 TAATGTGGGAAGTCTTGAGGACATATAATATGCGGTGAGAGAGAGATC 5935  
||| ||| |||  
872 ...ValGlnGlyArgGluAlaCysGluMetIleArgArgAlaAsnA 887  
5936 CGAATTTGGAAGGCTCAACAGTGTACACAGTTAACTCCAGGTTGGTA 5985  
||||| ||| |||  
887 spAsnValProGly.....SerAlaThrLysLysGluAsnValArgThrPhe 901  
5986 AACTTTATCGTCAGG.....GGAATGTATAAAAGGTTTGGT 6023  
||| ||| |||  
902 AspSerPheValMetAsnArgLysProGlyLysPheLysThrLeuTrpVa 918  
6024 GGATGAGTGCACATGATGATCAAGGCTTACTACAACTAGGCGTCTTCG 6073  
||||| |||  
918 lAspGluGlyLeuMetValHisProGlyLeuIleAsnPheCysIleAsnI 935  
6074 CAACCGCGCGTCGGAAGGCTCTTTTGGAGACATAAATCAGATACCA 6123  
::: ||| ||| |||  
935 leSerCysValSerValTyrIlePheClyAspArgLysGlnIlePro 951  
6124 TTCATAACACAGGAGAGGTGTTTAGCATGCTGCTGCTGCTTTTTCCTCC 6173  
||||| |||  
952 PheIleAsnArgValMetAsnPheSerIleProAspAsnLeuAlaLysLe 968  
|||  
6174 AAGAAGGAAGCGTTGTATACACTTCTAAATCGTACAGGTGCCGTTAG 6223

```
968 uTyrrAspGluIleValSerArgAspThrThrLysArgCysProLeuA 985
      :::::::::::::: :::: :::: :::: :::: :::: :::: ::::
6224 ATGTTTGCTACTGTGTGCTCCCA.....ATGACCGTA 6255
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
985 spValThrHisPheLeuAsnSerValTyrrGluLysArgValMetSerTyrr 1001
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
6256 AGGGGAACGGAAGAGTGTACCCCTGAAAGGTCGTTAGCGGTAAGGACAA 6305
      :::::::::::::: :::: :::: :::: :::: :::: ::::
1002 SerAsnValGlnArgSerLeuGluCysMetIleSerGlyLysAlaLys 1018
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
6306 ACCAGTACTAGATCCGTGTCCTCAAAAGGCCAATTGGAACCACTGATGACG 6355
      | :::::::::::::: :::: :::: :::: :::: :::: ::::
1018 s.....IleAsnAspT 1022
      :::::::::::::: :::: :::: :::: :::: :::: ::::
6356 TAGCTGAATAAAGCTGAC...GTGTACTGTGTCATGACCCAGTTGGAG 6402
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1022 yrArgSerIleLeuAlaGluGlyLysLeuLeuThrPheThrGlnGluAsp 1038
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
6403 AAGTCGGATATCAAGAGCTGCTTGAAGGGAAGGAAAGAACACACCACT 6452
      ||||| :::: :::: :::: :::: :::: :::: ::::
1039 LysGluTyrrLeu.....LeuLysAlaGlyPheLysAsp.....Va 1050
      :::: :::: :::: :::: :::: :::: ::::
6453 GATGACATGTCATGAAGCACAGGGAACAACTTCAGTGTGTTATGTT 6502
      | ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
1050 lAsnThrValHisGluAlaGlnGlyGluThrTyrrArgAspValAsnLeuI 1067
      :::: :::: :::: :::: :::: :::: ::::
6503 TTAGGACGAAGNAGCCATGACCTCCCTATTCACTAAA....CAACCGCAT 6549
      :::: :::: :::: :::: :::: :::: ::::
1067 leArgValThrAlaThrProLeuThrIleValSerAlaGlySerProHis 1083
      :::: :::: :::: :::: :::: :::: ::::
6550 ATACTGTGTTGTTGTCGAGACACACACGCTCACTGGTTTATGCGCTCT 6599
      :::: :::: :::: :::: :::: :::: ::::
1084 ValThrValAlaLeuSerArgHisThrAsnArgPheValTyrrThrVa 1100
      :::: :::: :::: :::: :::: :::: ::::
6600 GAGCTCAAGTTGGACGATACAGTCGGGCACATATATTAGCGACGCTCAC 6649
      :::: :::: :::: :::: :::: :::: ::::
1100 l.....ValProAspValValMetThrThrValGln.....LysT 1112
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
6650 CTCATCATGATCCGACGCTTGTCTTCACAGCTTCGCCCCCGCT 6693
      ||||| :::::::::::::: :::: :::: :::: :::: :::: ::::
1112 hrGlnCysValSerAsnPheLeuLeuAspMetTyrrAlaValAla 1126
      :::: :::: :::: :::: :::: :::: ::::
seq_name: SwissProt_39:V125_AMVLE
seq_documentation_block:
ID V125_AMVLE STANDARD: PRT: 1126 AA.
AC P03589;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE 125 KDA PROTEIN.
OS Alfalfa mosaic virus (strain 425 / isolate Leiden).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Alfamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-8314345; PubMed-6298738;
RA Cornelissen B.J.C., Brederode F.T., Moormann R.J.M., Bol J.F.;
RT "Complete nucleotide sequence of alfalfa mosaic virus RNA 1.";
RL Nucleic Acids Res. 11:1253-1265(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L00163; AAA46289.1; -.
DR PIR: A04197; WMFM12.
DR INTERPRO: IPR000606; -.
1948 .....
```

```
DR INTERPRO: IPR0002588; -.
DR PFAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Methyltransf_1.
SQ SEQUENCE 1126 AA; 125827 MW; BF5A8019B47D4CBF CRC64;

alignment_scores:
  Quality: 329.50      Length: 1884
  Ratio: 0.504        Gaps: 63
  Percent Similarity: 34.713      Percent Identity: 16.879

alignment_block:
US-09-301-906-3 x V125_AMVLE ..
Align seg 1/1 to: V125_AMVLE from: 1 to: 1126

1303 AGGAATTTGTCGATATGTTTCTCAA.....AGGA 1334
      ||| :::: :::: :::: :::: :::: ::::
40 ThrAlaIleGlyLysValPheSerGluAlaGlyArgAlaTyrrAlaGlnAs 56
1335 CGCAATGATTCGGAAGAACCTCACATAGGTGCGATGTTT..... 1377
      ||||| :::: :::: :::: :::: :::: ::::
56 PAlaLeu.....ProSerAspLysGlyGluValLeuLysIleS 69
1378 .....CTGAAGCCGCGGAAAGGAGAGCTCAGGCAACTCTTTCCA 1419
      ||| :::: :::: :::: :::: ::::
69 erPheSerLeuAspAlaThrGlnGlnAsnIleLeuArgAlaAsnPhePro 85
1420 GAGCTTTTCGATACAGTTCTCCGACTCGGTGAGGAGTAGTCACCCATTCG 1469
      ||||| :::: :::: :::: :::: ::::
86 GlyArgThrValPheSerAsnSerSerSerSerSerHisCysPheAl 102
1470 TAATGCCATGCGGACCTGTTCAATGGAATCTTTTCCAGGAGGTGTGTA 1519
      ||| ||| ||| :::: :::: ||| ||| |||
102 alaAlaHisArgLeuLeuGluThrAspPheValTyrrArgCysPheGlyA 119
1520 ATGTG.....TGCTTCTTCGATATTTGGGGGAGCTTCAGTCATCATGTC 1563
      ||||| :::: :::: :::: :::: :::: |||||
119 snThrValAspSerIleIleAspLeuGlyGlyAsnPheValSerHisMet 135
1564 AAAGTCGCCATGTGAACCTGTCATGTATGCAATCAGTCCTAGAGCTTAA 1613
      ||||| :::: ||| ||| ||||| ::::
136 LysValLysArgHisAsnValHisCysCysCysProIleLeuAspAlaAr 152
1614 AGATGTGAAGCGGAGACAATCAATGAGATCCTCTTTCTTCCACAGCTGGG 1663
      ||||| ||| :::: :::: ||||| |||||
152 gaspGlyAlaArgLeuThrGluArgIleLeuSerLeuLys..... 165
1664 GAGATTCGTACGTGTCAGT.....GACCTTCTTAACCTGAAGCGGCTTCA 1707
      ||||| ||||| :::: :::: :::: :::: |||||
166 .....SerTyrrValArgLysHisProGluIleValGlyGluAla..... 178
1708 AGTCTGTGCTTACTGTAGTCGAGAATCGCAACTCGGATCTGTAGAGC 1757
      :::: ||||| ||||| :::: |||||
179 .....AspTyrrCysMetAspThrPheGlnLysCysSerArgArgAl 192
1758 CGATCGCGGTTTATGGTG...GATGTGTACGATATATCCCGCAGCAGG 1804
      ||||| :::: :::: :::: :::: |||||
192 ansPtyrrAlaPheAlaIleHisSerThrSerAspLeuAspValGlyGluL 209
1805 TAGCAGAGCTATGATAAGAGGTCGCTGTTTTCACATAGCTCTT 1854
      ||||| :::: :::: :::: :::: |||||
209 euAlaCysSerLeuAspGlnLysGlyValMetLysPheIleCysThrMet 225
1855 ATGTTCCCGCTGGAGTGTGTACGGTAAC...GGTGAAGTTTACTTGA 1901
      ||| :::: :::: ||| |||||
226 MetValAspAlaAspMetLeuIleHisAsnGluGlyGluIleProAsnPh 242
1902 AGAATCTCATAGTGTGGTGAAGAGGAGGTGATTACCTGGGCTAC.... 1947
      :::: :::: :::: ||||| |||||
242 easnValArgTrpGluIleAspArgLysLysAspLeuIleHisPheAspP 259
1948 .....
```

```

259 heileAspGluProAsnLeuGly.....TyrSerHisArg 270
1984 TTCTCTAACCTAAGCGGGTTTTCACCTTTTCTATGTACGCACTTCGTC 2033
||||| : : : : :
271 PheSerLeuLeuLysHisTyrLeuThrTyr..... 280
2034 CGGGAACGCTTTTANGCTAGCATGAGGATACCGTTGGTTACCATC 2083
||||| : : : : :
281 ....AsnAlaValAspLeuGlyHisAlaAlaTyrArg..... 291
2084 ATCTCACTATGTAGGCTCAGAAAGTCACCTGGAAC.....GAG 2124
: : : : :
292 .....IleGluArgLysGlnAspPheGlyGlyValMetValIleAsp 305
: : : : :
2125 GTTACGTATAGTCGTTG.....GTCCCGTCGTCGTGG 2159
: : : : :
306 LeuThrTyr...SerLeuGlyPheValProLysMetProHisSerAsnG1 321
2160 CAAATCGCTGCTTCATACCTGTTAGCTGGTCTTAGTGTG..... 2202
||||| : : : : :
321 YArgSerCysAlaTrpTyrAsnArgValLysGlyGlnMetValValHist 338
2203 .....TCCTTTAAGACAATAGTCCTCGAT 2226
||||| : : : : :
338 hrValAsnGluGlyTyrTyrHisHisSerTyrGlnThrAlaValArgArg 354
2227 TCGGACTTTGTCGACAGATCTATTCTACGCGCTCAACACTATAGGGAC 2276
: : : : :
355 LysValLeuValAspLysLysValLeuThrArgValThrGluValAlaLaph 371
2277 ATTTCGAGATAGAACGTTTGAGTATGCCGTTGGGGGGTCAGTCGCAMA 2326
: : : : :
371 eArgGlnPheArgProAsnAlaAspAlaHisSerAlaIleGlnSerIleA 388
2327 AGACC.....CATGTCATTACAGGGAGTCGCGTT 2355
||||| : : : : :
388 laThrMetLeuSerSerSerThrAsnHisThrIleIleGlyGlyValThr 404
2356 GTCCACGACGAAGTTGATATTTCTCTGATGATATGTTGGGGTTTACTTGT 2405
: : : : :
405 LeuIleSerGlyLysProLeuSerProAspAspTyrIleProValAlaTh 421
2406 CGCTGTTATGGCTCAGCGCATTAAG..... 2430
: : : : :
421 rThrIleTyrTyrArgValLysLysLeuTyrAsnAlaIleProGluMetL 438
2431 .....GATAGGGCAAGAGTATTTCGCTCCTATAACTTTATAAAA 2469
||||| : : : : :
438 euserLeuLeuAspLysGlyGluArgLeuSerThrAspAlaValLeuLys 454
2470 GCGAGTGAGGGAGTCTCCCGGGTCTTCAAGCTCTTCTTCAGACCGT 2519
: : : : :
455 GlySerGluGlyProMet.....TrpTyrSe 463
2520 AGCGGATTGTTTTCAGCGAGCTCCGCTCTATGCTAAGCAATGGTGC 2569
||||| : : : : :
463 rGlyProThrPheLeuSerAlaLeuAsp.....LysValAsnValP 477
2570 ACGATAACTTCAACGTTTGGAGACGCTTATGCTATGCCACGAGCGTTC 2619
: : : : :
477 roGlyaspPhe...ValAlaLysAlaLeuLeuSerLeuProLys..... 490
2620 ATCCGTAAGTACCTGGGCTGTGTTGTATACCATTTCGACTTCTCGAGC 2669
490 ..... 490
2670 TTCAGACAGTTGGAGCTCAGGGTCCCTTTGATATTTCAAGGAGACCT 2719
490 ..... 490
2720 TCGGTAGAAACTGAAGATAGTCGCTTCGGCTCTTCTCTAGGGCTATC 2769
||| ||||| : : : : :

```

```

491 .....ArgAspLeuLysSer.....LeuPheSerArgSer.... 500
2770 GTGGAAGATTCAATTAAGTTCATGAAGCAATGAAGACAGAGATGGAAA 2819
||||| : : : : :
501 .....AlaThrSerHisSerG1 506
2820 ACCCTCCCAATTACTGAAGATTCTGTATATGCTTCATATATGGGGAACG 2869
: : : : :
506 uArgThrProValArgAspGlu..... 513
2870 TTTCTAACGTCCTCACTGTACGAGGCGAGCTCTCTTGGCGGTTCCAAAGCG 2919
||||| : : : : :
514 ..SerProIleArgCys..... 518
2920 ACCGTGGTTTTCGAGTGTCTTAAGGGTTTGGTAGCTCTGCGGGCTGGAC 2969
518 ..... 518
2970 GAAGGCCCTTTTCGGGCATTAGTCGTTCTTTCCACAGGTTCACTATTCT 3019
||||| : : : : :
519 .....ThrAspGlyValPheT 524
3020 ACGACCGCGTTTAACTGAAGATGAAGGCTTGATGCTCTGTCGCGCACA 3069
||| : : : : :
524 Yr..... 524
3070 GAGATGCTATAAACCTCACCGGTGGGCATACCTGGAGAGCTCGCGCTAGC 3119
||||| : : : : :
525 .....ProIleArgMetLeu..... 529
3120 TGTGAGCAAGCTCGTAGCTGGACGAAGAATTTTGGAGTGAAGTTTCCT 3169
||||| : : : : :
530 ....LeuLysCysLeuGlySerAspLys.....PheGluSerValThrI 543
3170 TAAATGACTTCACCACTTTCTGTTTGGCGAATTAAGGTGCTTATCGGGATA 3219
: : : : :
543 leThrAsp..... 545
3220 TTCGTGGCGTCTTTGGTGGCGGCCCAATTTCATGGAAGTATAGGCGCG 3269
545 ..... 545
3270 AATTGGCGTAAACGCTAGAAGGTACGGGGCAGTAGTTACGAACACTCTAA 3319
545 ..... 545
3320 GCTCGTTAAGTTTCAACAGCCGCGGTGTTTACGCGCTTTAACCTCTAGC 3369
||||| : : : : :
546 .....ProArgSerAsnThrGluThr 552
3370 ACAGTATCCGGTGGATCTTTAGTCGTCGAAGAGAGGTTTTCGTCGGCGGT 3419
||||| : : : : :
553 ThrValasp.....LeuTyrGlnSerPheGlnLysLysI1 564
3420 GACCGTCACTAGGGCGACCGTAGCTAAACGTCACGTCCTTACGCTTGC 3469
564 e..... 564
3470 TATCGTTTCTACCTCATACGCCATTTCGGCTGCAGTATCTTAGGCATT 3519
564 ..... 564
3520 TGGGCACATGCTCTTCCACGGCCTTAATGTTTCTTCTTGGTTTAGGGAC 3569
: : : : :
565 .....GluThrValPheSerPheIleLeuGlyLy 574
3570 ATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAAGTTTGGAGGCTTCT 3619
: : : : :
574 sileAspGlyPro..... 578
3620 CCAATAATTGGTCCGCTGTTCCCGAGGTGCTTTGGCGAGGAAGAGTCTC 3669
578 ..... 578

```

3670 ACCTCATTGTTACTCCCTATTATACGCTAGGGTATCTTTGATCATAAGGG 3719  
578 ..... 578  
3720 CTTGCTTAACGACACCATACCTCAACTGCTTACCTCCACCGGTAGAGG 3769  
579 .....SerProLeuIleS 583  
3770 GGAGGAATGTGTACGATGAGAGCTTAGG. ....TATTACGGGACTTT 3813  
583 crAspProValTyrPheGlnSerLeuGluAspValTyrAlaGluIrp 599  
3814 CACTATGACGAAGTGTGGTCCACTCTGGGACTCAGCATGAAGCGGTTC 3863  
599 ..... 599  
3864 CGGTGACGATAACGATGATTCACACTCTAGTGTCTCAAGCTATGATGTTG 3913  
600 .....HisGlnGlyAsnAlaIleAspAlaSerAsnTyr..... 610  
3914 TCACAAATGTGCGGACGTGGGGATTAGCACCAACGGGAAGTTACTGGT 3963  
610 ..... 610  
3964 GAAGAAGAGACCCATTACCTCGAAGCGTGCATATACACTTATGTCGAGA 4013  
610 ..... 610  
4014 AGAGTTGCCCGCTCGCAGTGTGGCGGAAGACAAGGTATCGTCGG 4063  
610 ..... 610  
4064 GTTCTGGTACCCTCAGCGTATGGCTTTGTGAAGTGTGNAAAAAGGT 4113  
611 .....AlaArgThrLeu 614  
4114 GTCCAGCATGCTTTCCACCAACAGTCTAGTGGGAAAGCGTCTGTGAGGT 4163  
615 LeuAspAspIleArgLysGlnLysGluGluSerLeuLysAlaLysAlaLys 631  
4164 TGAGTGGACGCCAAAGGGTGTCTCCAGAAAGCTGTCGGTGAAGCGC 4213  
631 sGluValGluAspAlaGlnLysLeuAsnArgAlaIleLeu..... 644  
4214 CGACACAAGAAAGGAGAGAGTGCAGATGTTAACACACGACCAACCGCG 4263  
644 ..... 644  
4264 GTCAACGAAGCGCAGAGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCC 4313  
644 ..... 644  
4314 ACAGGCTGATATTCCAAGGTCACCCAGTCCAGGTATACATGCTCAGAAAG 4363  
645 .....GlnValHisAlaTyrLeuG 651  
4364 AGTGAACAACAAGATACCATTTGGCGACTGTTTCGGCGCCACGCCAATC 4413  
651 LuAlaHisProAsp ..... 655  
4414 GTCGATGACAAACCCGCCCAAGTCTTACGACTCGTGGTGAAGATAAT 4463  
656 .....GlyGlyLysIleG1 660  
4464 TGACAAGGCAAGCGCGTCTCATGTGGCTGAGAAAAACAGGTACAAAG 4513  
660 uGlyLeuGlyLeuSerSerGlnPheIleAlaLysIleProGluLeuAlaI 677  
4514 TCGACGACCCCAACAGAGGAGTTTGACGATCAATGAAGGCAAGCCGGT 4563  
677 leProThrProLysProLeuProGluPheGluLysAsnAlaGluThrGly 693

4564 AAACAGCTTTGCATGTTTAGAACGTGTTCTCGGTGTGTCAGCTGGAGT 4613  
694 GluIleLeu.....ArgIleAsnPr 700  
4614 GTACAACGAGCGACTATGCCACCAGGTTCTCAAAACGCATTACCTTTG 4663  
700 ohisSerAspAlaIleLeu.....GluAlaIleAspTyrL 712  
4664 TCGATAACTTGAAGGAGGAGTGGGTCTTTCTCAAAGCTGGGTGAG 4713  
712 euLysSerThrSerAlaAsnSerIleIleThrLeuAsnLysLeuGly... 727  
4714 GGGTATACCTATAATGGTGTGATGCCCATGTTTCATCAGGGTGGCCTCGTGC 4763  
727 ..... 727  
4764 CCTAGAGGATATCTTAACGGCAATTAAGTACCAAGCGTCTTCGACCACT 4813  
728 .....AspHisC 730  
4814 GTTTAGTGCAGAAGTACAAGATGGGTGGAGCGTACCATTTCCACGCTGAT 4863  
730 ys.....GlnTrpThrThrLysGlyLeuAspValIrpAlaGlyasp 744  
4864 GACGAGGAGTGTCTATCCATCAGATAAACCCCTATCTTGACGGTCAATCTCCT 4913  
745 AspLys..... 746  
4914 GGGGAAGCAAACTTCTCGACTAAGTGCAGGAAGGTGTTAAGGTCATGG 4963  
746 ..... 746  
4964 TCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCTTGCCTTTCGTTTCAA 5013  
746 ..... 746  
5014 AGGACGCACCTTGCAATTCAGTAAACTCCATCGACGAAGGCGGCATCAGTTT 5063  
746 ..... 746  
5064 GACGTTACGGGCAACTCGGCGGTCTTTGGTGTAGCAGGATGTTGCAGT 5113  
747 .....ArgArgAlaPhe..... 750  
5114 TAGCCGGCGCGTGTGGATGAGAAGTCCACAGGTGTTCCAAACCCAGCAA 5163  
751 .....IleProLysLysasn 755  
5164 CCACAGAGCCAAAGGTGCTACCAAGAACAAATCACACCAAAATCGGGGGCAA 5213  
756 ThrTrpValGlyProThrAlaArgSerTyrProLeuAlaLysTyrGluAr 772  
5214 GGCTCTATCTGAGGAAGTGGTAGGAAGTCAAGGGGAGTGCACATACT 5263  
772 gAlaMetSer..... 775  
5264 CGATATGTTGCCAACAAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313  
776 .....LysAspGlyTyrValThr.....LeuArgTrp 784  
5314 GATAATCCAGTATGGCTCTTGAAGCTGACTACACCCCAATGACATTTGA 5363  
785 AspGlyGluValLeuAspAlaAsnCysValArgSerLeuSerGlnTyrG1 801  
5364 AGTGGTTAAACCGGG.....ACCTCTGAAGATGCCGTGC 5398  
801 uIleValPheValAspGlnSerCysValPheAlaSerAlaGluAlaIleI 818  
5399 TGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCG 5448  
818 leProSerLeuGlu.....LysAla 824  
5449 TTGCTTATGGCTAGAAATATTTCCGCTCACTACCGCGCGAAGGTGTTCTG... 5496



DR PFAM; PF01443; Viral\_helicase1; 1.  
KW PFAM; PF01660; Vmethyltransf; 1.  
FT Transforase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.  
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.  
FT NP\_BIND 833 840 ATP (POTENTIAL).  
SQ SEQUENCE 1616 AA; 183542 MW; 5DBFBF2FADCC5C0C CRC64;

alignment\_scores:  
Quality: 313.00 Length: 1879  
Ratio: 0.455 Gaps: 63  
Percent Similarity: 36.615 Percent Identity: 16.871

alignment\_block:

US-09-301-906-3 x RRPO\_TOMS1 ..

Align seq 1/1 to: RRPO\_TOMS1 from: 1 to: 1616

```
1207 GCCCTTAATCAAGAGGAGGTCTATTATCATCGGTTCGTTTCGAGTATCACCAA 1256
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 AlatyThrGlnThrAlaThrSerSerAlaLeuLeuGluThrIleArgG1 18

1257 TAGACTGCTATTAGAGGACCAATCGGCAATTCGTTTCCCATTTGGACACGA 1306
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

1307 AATTCTCGGATATGTTTCTCAAGGACCGCAATGATTCCGCGAAACCC 1356
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 laVal...Asp6luPheAsnAlaArgsp.....ArgArgPro 46

1357 TCACATAGTTCGGATGTGTTTCTGAAGCGGGGAAAGGAGAGCTGAG 1406
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

1407 GGAATCTTTCAGAGCTTCGATACAGTTCTCCGACTCGGTGAGGAGTA 1456
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlav 80

1457 CTCACCCATTCGTAATCCCATCGGAGCTCT...TTCAATGGAATCTTT 1503
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

1504 TCCAGGAGGTGTGTAATGTGTGTTCTTCGATATTGGGGGAGCTTCAC 1553
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

1554 GTATCATCTCAAGCTGCCATGTGAAGTGTATGTCATGTCGAATCCAGTCC 1603
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 aserHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

1604 TAGAGCTTAAAGATCTGAAGCGGAGAAATCAATGAG.....ATC 1641
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

1642 CTCCTTTCTTCC...ACAGCTGGGGGAGTTCGACGTGTCACGT..... 1683
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheG1 163

1684 .....GACCTTCTAACTGAAGCGGCTTCAAGTCTGTGTCTT 1720
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

1721 AC.....TGATGCGAGATCGCAGAACTGCGATCTAGA 1755
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 isAspThrPheGlnThrCysGlyHisSerGlnGluCysTyrThrGlyArg 196

1756 GCCGATCGGGTTTTATGTTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

1806 AGCAGAGGCTATGGATAGAGAGGTGCGCTGGTTTCGACATAGCTCTTA 1855
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
213 eelyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230
1856 TGTTCCTCCCGTGGAGTGTGTGTAACGGTGAAGTTTACTTTGAAGAA 1905
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
230 isPheSerGluAsnLeuLeuGluAspSerHisValAsnLeuAspGlu 246
1906 CTCGATAGTGTGGTGAAGAGGAGGTGATTACCTGGCGCTACAATCTTGG 1955
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263
1956 TCAGTGTGGTGAGATG...TATGAACATCTCTTCTACGTAAAGCGGT 2002
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 aserGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280
2003 TTTTCACCTTTTCTTATGTACGACTTCGTCGGGAACGTGTTTAAGCTA 2052
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
280 yrValCysLysThrTyrPheProAlaSerAsnArgIluValTyrMetLys 296
2053 GAGTATGAGGGATACCGTGTGTGTTACCATCATCTCACTATGTAGGGC 2102
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 GluPhe..... 298
2103 TCAGAAGTCACTGGAACCTGAGGTACGTATAGGTTCGTGGTCCCGTCT 2152
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 ..... 298
2153 TCGTGGCAATCGCTGGTGTTCATACCTGTGTAGCTGGTTCAGTGTG 2202
298 ..... 298
2203 TCCTTTAAGACAATAGTCTCTCGATTTCGGACTTTGTTCGACAGGATCTATTC 2252
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 .....LeuValThrArgValAsnTh 305
2253 CTAC.....GGCTCAACACTATAGGACATTCGAGAATAGAACGTTTG 2296
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 rTrpPheCysLysPheSerArgIleAspThrPhe..... 316
2297 AGTATGCCGTTTGGGCGGTTCAGTTCGCAAAAGACCATGTCATTACAGGG 2346
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 .....LeuLeuTyr 319
2347 AGTCGCGTTTGTCCACAGCAAGTTGATATTTCTCTGTGATATGTGGGG 2396
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 LysGlyValAlaHisLysGlyValasp..... 328
2397 TTTAGTTCTCGCTGTATGGCTCAGCGCATTAAGGATAGGGCGAAGAGTA 2446
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 ..... 328
2447 TTCGCTCTATAACTTTATAAAGCCAGTAGGGGAGTCTCCCGGGGTC 2496
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341
2497 TTCAGCTCTTCTTTCAGACCGTAGCGATTGTTTTTCGAACGAGTCTC 2546
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 .....TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355
2547 CGTCTATGCTAAGCAATGGTCACGATAACTTCAACGTTTTCGAGACGC 2596
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 uLeuGluAspSerSerValAsnTyrTrpPheProLysMetArgAspM 372
2597 TTATCTCTATGCCAGAGCGTTTCATCCGTAAAGTACCTGGGTCCTGTGT 2646
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 etValIleValPro..... 376
2647 GTTACCATTTGCATCTTCGAGCTTCAGACAGGTTTGAGCTCAGGGTGC 2696
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
377 .....Le 377
2697 CTTTCATATTTTCGAAGGAGACCTTCGGTAGGAACTGAAGAACTAGTCGT 2746
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
377 uPheAspIleSerLeuGluThr.....SerLysArgThrArgL 390
```



[illegible][illegible]



6286 GTCGTTAGCGGT.....AAGGACAACACAGTAGTAGAGTCGCTGCCAA 6329  
 DR PFAM; PF00978; RNA\_dep\_RNAPol2; 1.  
 DR PFAM; PF01443; Viral\_helicase1; 1.  
 999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGI 1015  
 KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
 6330 AAGGCCAATTTGAACCACTGATGACCTAGCTGAATAAAGCGTGACGTGT 6379  
 FT CHAIN 1 1616  
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.  
 FT NP\_BIND 833 840 ATP (POTENTIAL).  
 1015 yLys.....  
 6380 ACTTGTGCATGACCCAGTGGAGAACTCGATATGAAGAGGTGGTTGAAG 6429  
 1017 leLeuThrPheThrGln.....SerAsp...LysGluAlaLeuLeu 1029  
 6430 GGAAGAGGAAAGACACACAGCTGATGACAGTCATGAAGCAGCAGGAA 6479  
 1030 SerArgGly...TyrThrAspValHisThrValHisGluValGlnGlyGI 1045  
 6480 AACATTCAGTGTGTGTATTTAGGACGAGAAAGCGGATGACTCCC 6529  
 1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062  
 6530 TATTTACTAAA...CAACCCGATATACTTGTGGTTTGTGACGACACACA 6576  
 1062 leilealaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078  
 6577 CGCTCACTGGTTATGCCGCT.....CTGAGCTCAAAGTT 6611  
 1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleI 1095  
 6612 GGACGAT.....AAGTCGGCACATATATAGCGAC 6642  
 1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq\_name: SwissProt\_39:RRPO\_TOML

seq\_documentation\_block:  
 ID RRPO\_TOML STANDARD; PRT; 1616 AA.  
 AC P03587; O41352;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE RNA-DIRECTED RNA POLYMERASE (BC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:  
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].  
 OS Tomato mosaic virus (strain L) (TMV) (TMV strain tomato).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85157522; PubMed=6549393;  
 RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,  
 RA Okada Y.;  
 RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)  
 RT genome and comparison with the common strain genome.";  
 RL J. Biochem. 96:1915-1923(1984).  
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL  
 CC RNA REPLICATION  
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA  
 CC CAPPING AND AN RNA HELICASE.  
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS  
 CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X02144; CAA26085.1; --  
 DR EMBL; X02144; CAA26082.1; --  
 DR PIR; A04195; WMTM8T.  
 DR INTERPRO; IPR000606; --  
 DR INTERPRO; IPR001788; --

DR INTERPRO; IPR002588; --  
 DR PFAM; PF00978; RNA\_dep\_RNAPol2; 1.  
 DR PFAM; PF01443; Viral\_helicase1; 1.  
 KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
 FT CHAIN 1 1616  
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.  
 FT NP\_BIND 833 840 ATP (POTENTIAL).  
 SQ SEQUENCE 1616 AA; 183564 MW; A8EC8929B5CF7CAF CRC64;

alignment\_scores:  
 Quality: 309.00 Length: 1879  
 Ratio: 0.445 Gaps: 64  
 Percent Similarity: 36.935 Percent Identity: 16.764  
 alignment\_block:  
 US-09-301-906-3 x RRPO\_TOML ..  
 Align seg 1/1 to: RRPO\_TOML from: 1 to: 1616

1207 GCCTTAAATCAAGAGGAGGTCTATTTCATCGCTGCTTTCGAGTATCACAA 1256  
 2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGI 18  
 1257 TAGACTGGTATTAAAGGACCAATCGGCATTGCTTCCCATTTGGACACGA 1306  
 18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgLeuTyrAspThrA 35  
 1307 AATTGTGCGATATGTTTCTCAAGGGAGCCCAATGATTCGCGAANAACCC 1356  
 35 laVal...AspGluPheAsnAlaArgasp.....ArgArgpro 46  
 1357 TCACATAGTGGCATGTGTTTCTCAAGCCCGGGAGGAGGAGGAGCTGAG 1406  
 47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63  
 1407 GGAACCTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGTCAGGAGTA 1456  
 63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlav 80  
 1457 GTCACCCATTCGCTAATGCGATCGCGAGGTGT...TTCATGAAATCTTT 1503  
 80 aHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96  
 1504 TCACAGAGGTGTGTAATGTGTCTTCTCGATATATTCGGGGAGCTTCAC 1553  
 97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113  
 1554 GTATCATGTCAAAGCTGGCCATGTGAACCTGTATGATGATGCAATCCAGTCC 1603  
 113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130  
 1604 TAGACCTTAAAGATGTCAAGCGGAGATCAATGAG.....ATC 1641  
 130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146  
 1642 CTCCTTTCTTTCC...ACAGTCGGGAGATTCCTACCTGTGCC..... 1680  
 147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGI 163  
 1681 .....AGTGACCTTCTAAGTGAAGCGGCTTCAA 1708  
 163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180  
 1709 AGTCTGTGTTACTACTAGTCGAGATCGCAGAACTGC...GATTCTAGA 1755  
 180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196  
 1756 GCGGATCGCGGTTTATGCTGGATGTGTACGATATATATCCCGCAGCAGGT 1805  
 197 valTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213





```
5396 TCGTCGAGTACTTGAAGTATCTGGCTATAGGCATTTGAGAGACATACAGG 5445
|||||
801 alVal.....TyrSerAspMetAlaLysLeuArgThrLeuArg 813
5446 GCGTTGCTTATGGCTAGAAATATTCGCCGTCACCTACCGCGGAGGTGTTCT 5495
|||||
814 ArgLeuLeuLysAspGlyCylProHisValSerSerAlaLysValValLe 830
5496 GAAAGTACCTAATCAAGTTTATGAATCATCTACCGGGCTTTCCACGTTTACA 5545
830 |
830 u..... 830
5546 AGTCGGGCACAGATCTCATCTTTTTCATTCAACACAGACGGCTTGCCTGTG 5595
830 ..... 830
5596 AGAGACCTACGCTACGCTACTCATAGCTCAAAAGATATCTTTACCAAGGG 5645
830 ..... 830
5646 CAAAGATGTCGACGGGTGGTAGCTTTGGCGCACAATCTGTTGCTATCGG 5695
831 .....ValaspGlyVal..... 834
5696 ACGATATACTGGTTTTCACGATGCCATTAATTTGATAGTGCACGTGAAA 5745
834 ..... 834
5746 GTCGCTCATGGCGCATCGTGGCGCAATCGTTAAGTCCCTTCGATATAA 5795
834 ..... 834
5796 GTGCTATAATGCTCCCGCAGGTGGGGTAAGACGACGACGTTAGTGGACG 5845
835 .....ProGlyCysGlyLys..... 839
5846 AATTCGTTTAAGTCAACCAATAGCACAGCCACCATTCACGGCTAATGTGGGA 5895
840 .....ThrLysGluIleLeuSerArgValasn 848
5896 AGTCTGAGGACATAAATATGGCG.....GTGAA 5924
849 PheGluGluAspLeuIleuValProGlyArgGlnAlaAlaGluMetIle 865
5925 GAAGAGAGATCCGATTTGGGAAGGTCTCAACAGCTGCTACACAGTTAACT 5974
865 eaRgArgArgAlaasnAlaSerGlyIleIleValAlaThrLysaspasnV 882
5975 CCAGGCTGCTA...AACTTTATCGTCT.....AGGGGA 6003
882 alArgThrValaspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898
6004 ATGTATAAAGGGTTTTCGTGGATGAGTGCACATGATGCATCAGAGCTT 6053
899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyC 915
6054 ACTACAAGTACGCGTCTCGCAACCGCGCTCGGAGGCGCTCTTTTGTG 6103
915 sValasnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932
6104 CAGACATAAATCAGATACCATTCATAAAGGAGAGAGGTGTTTAGTAGTG 6153
932 lyAspThrGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948
6154 GATTCGTGCTGTTTGTTCACAAAGAGAAAGCGTTGTATACACTCTCAA 6203
949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965
6204 ATCCATACAGGTGTCGTTAGTGTGCTACTTGTGTCTCTCA..... 6246
965 rThrLeuArgCysProAlaaspValThrHisPheLeuasnGlnArgTyrG 982
6247 .....ATGACCGTAAGGGGAACCGAAAGTGTACCCTGAAAAG 6285
```

```
982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998
6286 GTCGTTAGCGGT.....AAGGACAAACAGTAGTAAGATCGCTCTCCAA 6329
999 MetValSerGlyAlaAlaSerIleasnProValSerLysProLeuLysGl 1015
6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACCGTCACGTCT 6379
1015 yLys.....I 1017
6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
1017 leLeuThrPheThrGln.....SerAsp...LysGluAlaLeuLeu 1029
6430 GGAAGAAGGAAGAAACACCACTGATGACGTGACATGAAGACACAGGAAA 6479
1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyGl 1045
6480 AACATTCACTGATGCTGTTATGTTTAGGACGAGAAAGCCGATGACTCCC 6529
1045 uThrTyrAlaaspValSerLeuValArgLeuThrProThrProValSerI 1062
6530 TATTCACATAAA...CAACCGCATATCTGTTGTTGTCGAGACACACA 6576
1062 leIleAlaArgAspSerProHisValValLeuValSerLeuSerArgHisThr 1078
6577 CGCTCACTGCTTTATGCGCGCT.....CTGAGCTCAAGTT 6611
1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIle..I 1095
6612 GGACGAT.....AAGGTGCGCACATATATTAGCGAC 6642
1095 eaRgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107
seq_name: SwissProt_39:RRPO_TOMK2
seq_documentation_block:
ID RRPO_TOMK2 STANDARD; PRT; 1616 AA.
AC P99676; P90349;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain Kazakh K2) (TOMV) (TMV strain K2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98012318; PubMed-9454068;
RA Belenovich E.V., Genetozov E.V., Novikov V.K., Zavrilev S.K.;
RT "Properties and structure of the tobacco mosaic virus strain K2
genome.";
RL Mol. Biol. (Mosk) 31:826-830(1997).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z92909; CAB07439.1; -
CC EMBL; Z92909; CAB07438.1; -
CC INTERPRO; IPR000606; -
CC INTERPRO; IPR001788; -
```



|||||  
377 uPheAspIleSerLeuGluThr.....SerLysArgThrArgL 390  
2747 TCGCGCTCTCTAGGCTATCGTGGAGATTCAATT...AAGTCATG 2793  
390 ysGluValLeu.....ValSerLysAspPheValThrValLeu 403  
2794 AAGCAATCAACACAGAAGATCGAAACCCCTGCCAATTACTGAAGATTC 2843  
404 AsnHsIleArgThrTyrGlnAlaLysAlaLeu.....ThrTyrSerAs 418  
2844 TCTATATCGGTTCAATAATGGGCAAGCTTCTAACTCCACTGCTAGAGGG 2893  
418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430  
2894 CAGGCTCTCTGGCGGTTCGAAAGCGCGTTCGAGTGTTCCTAAG 2943  
431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445  
2944 GGTTCGTAGTCTGTCGGGCTCCGACGAGGCGCTTTCTGCGATTACGTC 2993  
446 SerLeuLeuGlnSerLeuSerMetThr..... 454  
2994 GTTCCTTTCCACAGTTCACATTCTACAGCGCGGTTAACCTGAAGATG 3043  
455 .PhePhe..... 456  
3044 AAGGCTTGATGCTCTGTCGGCGCACAGAGATGCTATAAATCACCCTG 3093  
457 .....LeuHsThrLysLeu 461  
3094 GGCATCTGGACACCTCGCGGTAGCTGTGACGAAGTCGTAGCTGGAAC 3143  
462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477  
3144 GAAA.....GAATTTTGGAGTGAAGTTTCCTTAATGACTTCA 3181  
477 olySthrValSerGlnHisValTrpAspGluIleSerLeu..... 490  
3182 CCACCTTCGTATTCGGGAATAGGTGCTATTCGGGATATTCGTGCGCT 3231  
490 ..... 490  
3232 TTGGTGGCGGCCCAATTCATGGAAGTATAGCGCGGAATTGCGCTAA 3281  
490 ..... 490  
3282 CGCTAGAAGGTACGCGGCGAGTAGTTACGAACTAAGCTGCTTAAGTT 3331  
490 ..... 490  
3332 CACAAGCCGCGGTGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381  
490 ..... 490  
3382 CGATCTTTAGTCGTCGAAGAGGTTTCGTGCGGCGTGAACGTCACCTAG 3431  
491 .....AlaPheGlyAsnAlaPheProSerIleLys 500  
3432 GCGCAACCTAGCTAAACGTCAAGTCCCTTACGCTGCTATCGTTTCTTA 3481  
500 sGluArgLeuIleAsnArgLys..... 507  
3482 CCTCATACGCCATTTCCGGCTCGAGTATGTTAGGCATTTGGGCACATGCT 3531  
508 .....LeuIleLysIleThrGluAsnAla 515  
3532 CTTCACGCGCACTTAATGTTTTCTTTGTTTAGGACATTCCTGGGCG 3581  
516 LeuGlu..... 517  
3582 GAGGCGTAGCGGAATACTTGGAAAGTTTGGAGGCTTCTCCAATAATGGT 3631

518 ..... 518  
3632 GGCCTGTTCCCGAG.....GTTGTTTGGCAGGGAAGAGTGTACAGTCA 3675  
518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534  
3676 TTGTTACTCCCTATTACGCTAGGGGTATCTTTGATCATAAAGGGCTTGC 3725  
535 TyrLysMetSerValAspMet..... 541  
3726 TAACGACACCATACCTCAACTTGTTCCTCCACCGGTAGAGGGAGGA 3775  
542 .....ProValLeuAspIleArgL 548  
3776 ATGCTACGATGAGACGCTTAGGTATTACCGGACTTTTCACTATGACGAA 3825  
548 ysLysMetGluGluThrGluGluMetTyr..... 557  
3826 GGTGCTGTCTCCATCTGGGACTCAGCATGAAGCGTTCGCGTGACGATAA 3875  
557 ..... 557  
3876 CGATGATCCACTTCTAGTGTCTCAAGTATGATGTTGTCAAAATGTGC 3925  
558 .....AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570  
3926 GCGACGTGGGATTAGCACCAACGGGAAGTTACTGTTCAAGAAAGAGACC 3975  
570 spLysPheAspValAspValPheSerGlnMetCys..... 581  
3976 CATTCACCTCGAAGCTGCAATACACTTATGTCGAGGAAGAGTTGCCCC 4025  
582 .....GlnSerLeu.....GluValAspPr 588  
4026 GTCTCAGCTGTGGCGGAAAGACAGGTATCCGTCGGTCTCTGTACCG 4075  
588 oMetThrAla..... 592  
4076 CTGACGCTATGCTTTTGTGAAAGTGTCAAAAAGGTGTCGACGATGTC 4125  
592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuTr 608  
4126 TTTCACCAACAGTCTAGTGGGAAACGGCTCGTGGTTTGGGTGACGG 4175  
609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu..... 621  
4176 CAAAGGTTGCTCCACAGAAAGCGTCGTCGTGAGGCGCGGCACACAGAAA 4225  
622 .....GlnAspS 624  
4226 GGGGAAGAGCTGCAGATGGTAACACAGCACAAACCCGCGTCAACGAAGC 4275  
624 erGluLysAlaSerAspGly..... 630  
4276 CACAGGAGCCAGTACAGTCCAGTCTTGTAGTTGCGCCACAGGCTCATAT 4325  
631 .....AlaLeuValValThrSerArgAspValGlucl 641  
4326 TCCAAAGGTCAACCGTCCGAGGTACATGCTCAGAAAGAAAGTGAACAAG 4375  
641 uProSerIleArgGlySer.....MetAlaArgGlyG 652  
4376 AAGTACCATTTGGGACTGTTTCGGGCGCCACCCCAATCGTCGATGAGAAA 4425  
652 luLeuGlnLeuAlaGlyLeuSerGly.....AspVal 662  
4426 CCCGCCCAAGTGTTCACGACTCGTGGTGTGAAGATAATTGACAAGGGCAA 4475  
663 ProGluSerSerTyrThr..... 668  
4476 GGGCGCTCATGTGGCTGAGAAAAACAGGTCAAGTCGAGCAGCCCA 4525  
669 .....ArgSerGluGluIleGluSerLeuG 677



4526 AACAGAGGAGTTTTCAGCATCAATGAAGGCAAGCCGGTAAACAGCTTTGCC 4575  
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
677 luGlnPheHisMetAlaThrAlaSerSerLeuLeuHisLysGlnMetCys 693  
4576 ATGTTTAGAAGCTGTTCCCTCGCGTGTCCAGCTGGATGTGTACAAACGAAGC 4625  
:||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
694 Serile.....ValThrThrGlyPr 700  
4626 GACTATCGCCACCAGGTCTTCAAACGCATTTACCTTTGTGCGATAACTTGA 4675  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
700 oLeuLysValGlnInLnmetylsasn.....PheIleAspSerLeuV 714  
4676 AAGGAGGAGTGGCGTCTTTTCTCAAAGCTGGGTGAGGGGTATACCTAT 4725  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
714 alAlaSerLeuSerAla..... 719  
4726 AATGGTGGTAGCCATGTTTCATCAGGGTGGCTCTGTCGCCCTAGAGGATAT 4775  
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
720 .....AlaValSerAsnLe 724  
4776 CTTAAGCGCAATTAAGTACCAGAGCGTCTTCGACCACCTGTTTAGTCGAGA 4825  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
724 uValLysIleLeuLysAspThrAlaAlaIleAsp.....LeuGluThrA 739  
4826 AGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATCAGCAGGAGTGC 4875  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
739 rgGlnLysPheGly..... 743  
4876 TATCCATCAGATAAACCTATCTTTCACGGTCAATCTCGTGGGGAAGCAAA 4925  
743 ..... 743  
4926 CTTTCGACTAAGTCGAGGAAGGTGGTAAGTTCATGGTCATAAACGTAG 4975  
744 .....ValLeuAspValA 748  
4976 CTTTCGGGTGACTATTTTCATTATGCCCTTGC.....GGT 5007  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
748 leSerLysArgTrpLeuValLysProSerAlaLeLysAsnHisAlaTrpGly 764  
5008 TTTCAAGGAGCAGCTTGCATTTCAGTAAAC.....TCCATCGA 5045  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
765 valValGluThrHisAlaArgIstYrHisValAlaLeuLeuGluHisAs 781  
5046 CGAAGGCGGCATCAGTTTGACGTTTCAGGCAACTCGCGCGCTTTTGGTG 5095  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
781 pGlu.....PheGlyI 785  
5096 TAGGCAAGGATGTTCCAGTTAGCGCGCGCGGTGTCGGATCAGAGTCAACA 5145  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
785 le..... 785  
5146 GGTGTTCCAAACCAGCAACCAACAGACCAAGGTGCTACCAGAACATCAC 5195  
785 ..... 785  
5196 ACCAAATCGGGGGCAAGGCTCTATCTAGGGAAGTGTAGGGAAGTCA 5245  
785 ..... 785  
5246 AGGGGAGTCGACATCTCGATATGTCGCAACAGATTACGTTAGGAAG 5295  
786 .....IleThrCysAsp..... 789  
5296 TGTGAGTGGCTCAGGCTCATATCCAGTCATGCGCTCTTGAACCTGACTA 5345  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
790 ...AsnTrpArg..... 793  
5346 CACCCCAATGACATTTGAAGTGGTAAAAACCGGACCTCTGTAAGATGCCG 5395  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
794 .....ValAlaValSerSerGluServ 801

5396 TCCTGCGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGCAGACATACAGG 5445  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
801 alVal.....TyrSerAspMetAlaLysLeuArgThrLeuArg 813  
5446 GCCTTGTCTTATGGCTAGAAATATTGGCTCAGTACCAGCCGAAGGTGTCT 5495  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
814 ArgLeuLeuLysaspGlyGluProHisValSerSerAlaLysValValle 830  
5496 GAAAGTACCTAATCAAGTTTATGAATCACTACCCGGGCTTTTACAGCTTTACA 5545  
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
830 u..... 830  
5546 AGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTCCGTGTG 5595  
830 ..... 830  
5596 AGAGACCTACCGTACGTACTCATAGCTGAAAGAGGTATCTTTACCAAGGG 5645  
830 ..... 830  
5646 CAAAGATGTGACGCGGTGGTAGCTTTTGGCGACAATCTCTTCGTATGCG 5695  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
831 .....ValAspGlyVal..... 834  
5696 ACCATATACTGGTGTTCACACGATGCCATTAAATTGTAGGTGCACATGAAA 5745  
834 ..... 834  
5746 GTCGCTCGATGCGGCATGTTGGCGGAATCGTTTAAAGTCTTCGAATATAA 5795  
834 ..... 834  
5796 GTGCTATAATGCTCCGCCAGTGGCGGTAAAGACGACGACGCTTAGTGGAGC 5845  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
835 .....ProGlyCysGlyLys..... 839  
5846 AATTCCTTAAGTCACCCATAGCACAGCCACCACCATTTACGGCTAATGTGGGA 5895  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
840 .....ThrLysGluIleLeuSerArgValAsn 848  
5896 AGTTCGTAGGACATAAATATGGCG.....GTGAA 5924  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
849 PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetII 865  
5925 GAAGAGAGATCCGAATTTGGAAGTCTCAACAGTGTCTACACAGTAACT 5974  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
865 eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882  
5975 CCAGGTGGTA...AATTTATCGTC.....AGGGGA 6003  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
882 aArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898  
6004 ATGTATAAAAGGGTTTTTGTGGATAGGTGCACATGATGCATCAAGCCTT 6053  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyCy 915  
6054 ACTTAACTAGCGCTCTTCGCAACCGCGCTCGGAAGGCGCTCTTTTGTG 6103  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
915 sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932  
6104 GAGACATAAATCAGATACCATTCATAACAGAGGAGAGGTGTTTGTAGATG 6153  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
932 lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948  
6154 GATTGTGCTGTTTTTGTTCARAGAGGAAGCGTTGTATACACTTCTAA 6203  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgth 965  
6204 ATCGTACAGGTGTCGGTTAGATGTTTGGCTACTTGTGTCTCTCA..... 6246  
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982  
6247 .....ATGACCGTAAAGGGGAACGGAAAGGTGTTACCGCTGAAAG 6285





```
491 .....AlaPheGlyAsnAlaPheP 497
3632 GCGCTGTTCCGAGGTCTGTTGGCGAGGAAGAGTGTCCAGCTCATTTGTTA 3681
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
497 roSerIleLysGlu.....ArgLeuIleAsnArgLysLeu 508
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3682 CTGCCTATTACCTAGGCGTATCTTTGTATCATCAAGGGCTTGCCTTAACGA 3731
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
509 IleLysIleThr.....GluAsnAlaLeuGluI 518
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3732 CACCATACCTCAACTGTTACGTCCCAACCGGTAGAGGGAGGAATGTGT 3781
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
518 eArgValProAspLeu...TyrVal..... 525
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3782 ACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCT 3831
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
525 ..... 525
3832 GGTCCATCTGGGACTCAGCATGAAGCGGTTCCTCCGTGACGATAACGATGG 3881
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
526 .....ThrPheHisArgLeu..... 531
3882 ATCCACTCTAGTGTCTCAAGCTATGATGTTGTCAAAATGTGCGCGACG 3931
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
532 *.ValSerGluTyrLysMetSerValAspMetProValL 544
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3932 TGGGATTAGCACCAACGGGAAGTTACTGTTGAAGAGACACCCATTCA 3981
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
544 euAspIleArgLysLysMetGluGluThrGluGlu..... 555
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3982 CCTCGAAGCGTCCAATACACTATGTCGAGGAAGAGGTTGCCCGCTCTGC 4031
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
556 .....MetTyrAsnAlaLeuSerGlu..... 562
4032 AGCTGTGCGGGAAGACAAAGTGATCCCTCGGTTCTCGTACCGCTGACG 4081
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
562 ..... 562
4082 CTATGGCTTTTGTCAAGTGTGAANAAGGTGTCGAGATGCTTTCAC 4131
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
563 ..LeuSerValIleLysThrSerAspLysPheAspValAspValPheSer 578
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4132 CAACAGCTAGTGGGGAAACCGCTGCTGAGGTTGAGTGGAGCGCAAGG 4181
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
579 GlnMet.....CysInSerLeuGluValAsp.....Pr 588
4182 GTTGCTCCCAAGACGTGCTGCGGTAGGCGCGCACACAGAAAGGGA. 4230
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
588 oMetThrAlaAlaLysValIleValAlaValMetSerAsnGluSerGlyL 605
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4231 .....AGAGCTGCAGATGGTAACACAGACACAAACCGCG 4263
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
605 euThrLeuThrPheGluGlnProThrGluAlaAsnValAla...LeuAla 620
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4264 CTCACGAAGCGCAGCAGCGCCAGTACAGTCCAGTCTGTGTGAGTTCCGC 4313
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
621 LeuGlnAspSerGluLysAlaSerAspGlyAlaLeuValValThrSerAr 637
4314 ACAGGCTGATATTCAAAGGTCAACCGTCCAGGTACGAGGTACATGCTCAGAAAG 4363
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
637 gAspValGluGluProSerIleArgGlySer.....M 648
4364 AAGTGNACNAGAGTAGTACCATTTGCGGCTGTTTCGGGCGCCACGCCAATC 4413
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
648 etAlaArgGlyGluGlnLeuAlaGlyLeuSerGly..... 660
4414 GTCGATGAGAAACCGCCCAAGTGTACGAGCTCGTGTGTGAAGATAAT 4463
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
661 .....AspValProGluSerTyrThr..... 668
4464 TGACAAGGGCAAGCGCGTCTGCTCATGTGGCTGAGAAAAACAGGTACAAAG 4513
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
669 .....ArgSerGluGluI 673
```

```
4514 TCAGCAGCCCAACACAGAGAGTTTTCAGCATCAATGAAGCAAGCCGGC' 4563
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
673 leGluSerLeuGluGlnPheHisMetAlaThrAlaSerSerLeuIleHis 689
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4564 AAACAGCTTTTCATGTTTAGAACGTGTTTCCTGCGGTGTGCAGCTGGATGT 4613
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
690 LysGlnMetCysSerIle.....Va 696
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4614 GTACAAACGAAGGACTATCGCCACCAGGTTCCTCAACGCAATTTACCTTTG 4663
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
696 ltyrThrGlyProLeuLysValGlnGlnMetLysAsn.....PheI 710
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4664 TCGATAACTTGAAGGGAGGAGTGCAGTCTTTTCTCAAAGCTGGGTGAG 4713
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
710 leAspSerLeuValAlaSerLeuSerAla..... 719
4714 GGGTATACCTATAATGGTGTAGCCATGTTTCATCAGGGTGGCCTCGTGC 4763
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
720 .....Al 720
4764 CTAGAGGATATCTTAACGGCAATTAAGTACCAAGCGTCTTCGACCACT 4813
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
720 aValSerAsnLeuValLysIleLeuLysAspThrAlaAlaIleAsp.... 735
4814 GTTTAGTGCAGAAGTACAAGATGGGTGGAGCGGTACCATTTCCACGCTGAT 4863
      ::::  |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
736 ..LeuGluThrArgGlnLysPheGly..... 743
4864 GACGAGGAGTGCTATCCATCAGATAACCCCTATCTTTCAGCGTCAATCTGT 4913
743 ..... 743
4914 GGGGAAGCAAACTTCTCGACTAAGTCAGGAAGGTTGTAAGGTCATGG 4963
744 .....V 744
4964 TCATAACGCTAGCTTCGGTGACTATTTTCTTATCGCTTGC..... 5004
744 alLeuAspValAlaSerLysArgTrpLeuValLysProSerAlaLysAsn 760
5005 .....GGTTTTCAAAAGCAGCAGCTTGCATTTCAGTAAAC..... 5037
761 HisAlaTrpGlyValValGluThrHisAlaArgLysTyrHisValAlaIle 777
5038 ....TCCATCGACGAAGGGCGCATCAGTTTGACGTTTCAGGGCAACTCGGC 5083
777 uLeuGluHisAspGlu..... 782
5084 GCGTCTTTGGTGTAGGCGAGGATGTTGCAGTTAGCGCGCGGTGTCGGAT 5133
783 .....PheGlyIle..... 785
5134 GAGAAAGTCACAGGTGTTCCAAACCAAGCAACACAGAGCCAAAGTCTAC 5183
785 ..... 785
5184 CAGAAATATCACCAAAATCGGGGGCAAGGCTCTATCTAGAGGAAGTG 5233
785 ..... 785
5234 GTAGGGAAGTCAAGGGAGGTGCGACATACTCGATATGTTGCAACAAGAT 5283
786 .....IleThrCysasp..... 789
5284 TACGTTAGGAAGTGTAGTGGCTCAGGCTGATAATCCAGTATGGCTCT 5333
790 .....AsnTrpArgArg..... 793
5334 TGAACCTGACTACACCCCAATGACATTTGAAGTGTAAACCGGACCT 5383
794 .....ValAlaValIs 797
```

```

5384 CTGAAGTCCGCTGAGTACTGTAAGTATCTGCGTATAGCATTTAG 5433
|||||
797 erSerGluSerValVal.....TyrSerAspMetAlaLysLeu 809
|||||
5434 AGCAGATACAGGCGCTGCTTATGCTGAGAAATATTGCGCTACTACCGC 5483
||||| ||| |||||
810 ArgThrLeuArgArgLeuLeuLysAspGlyGluProHisValSerAl 826
|||||
5484 CGAAGTGTCTCGAAAGTACCTAATCAAGTTTATGAATCACTACCGGCT 5533
|||||
826 aLysValValLeu..... 830
5534 TTCACGTTTACAGTCGGGCACAGATCTCATTTTTCATTCAACACAGAC 5583
830 ..... 830
5584 GCCTTCGCTGTGAGACACCTACCGTACGTACTCATAGCTCAAAAAGGTAT 5633
830 ..... 830
5634 CTTTACCAAGGCAAGATGTCACCGCTGTAGCTTTGGGCGCAATC 5683
|||||
831 .....ValAspGlyVal..... 834
5684 TGTTCGTATCGGAGATATCTGTTTTCACCATGCCATTATTTGATA 5733
834 ..... 834
5734 GTGTCACCTGAAAGTCGCTCGATCGCGCATGTTGGGGAATCGTTTAAGTC 5783
834 ..... 834
5784 CTTTGAATATAAGTGTATTAATGCTCCCGAGTGGCGTAAGACACGCA 5833
|||||
835 .....ProGlyCysGlyLys..... 839
5834 CGTTAGTGACGAAATTCGTTAAGTACCACTAGCACACCCACCATTCAG 5883
|||||
840 .....ThrLysGluLeuLeu 844
5884 GCTAATGTGGAGAGTCTCAGGACATAAATATGGCG..... 5919
|||||
845 SerArgValAsnPheGluGluSerLeuLeuValProGlyArgGlnAl 861
5920 .....GTGAAGACAGAGATCCGAATTTGGAAGTCTCAACAGTGCTA 5962
||||| ||| |||||
861 aLaGluMetIleArgArgAlaAsnAlaSerGlyIleIleValAlaT 878
5963 CCACAGTTAACTCCAGGNGGTA...AAGTTTATCGTC..... 5997
|| ||| |||||
878 hrLysAspAsnValArgThrValAspSerPheLeuMetAsnTyrGlyLys 894
5998 .....AGGGGAATGTATAAAGGGTTTGGTGGATGAGGTGCGACATGAT 6041
||| |||||
895 GlyAlaArgCysGlnPheLysArgLeuPheIleAspGluGlyLeuMetIle 911
6042 GCATCAAGCTTACTACAACTAGGCGTCTTCGCAACCCGCGCTCGGAAG 6091
||||| ||| |||||
911 uHisThrGlyCysValAsnPheLeuValGluMetSerLeuCysAspIleA 928
6092 GCCTCTTTTGGAGACATAAATCAGATACCATTCATAACAGGGGAGAG 6141
||| |||||
928 latYrValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValThr 944
6142 GTGTTTAGCATGTGCTGCTGTTTGTTCCAAGAAGAACGCTTCT 6191
||| |||
945 GlyPheProTyrProAlaHisPheAlaLysLeuGluValAspGluValG 961
6192 ATACACTTCTAAATCGTACAGGTGTCGGTTAGATGTTTCTACTGTTCT 6241
||| |||||
961 uThrArgArgThrThrLeuArgCysProAlaAspValThrHisPheLeuA 978
6242 CCTCA.....ATGACCGTAAGGGGACGGAAGAAAGTGT 6273

```

```

978 snGlnArgTyrGluGlyHisValMetCysThrSerSerGluLysLysSer 994
|||||
6274 TACCCTCAAAAGGTTCGTAGCGT.....AAGCACAAACACGAGTAGTAAG 6317
|||||
995 ValSerGlnGluMetValSerGlyAlaAlaSerIleAsnProValSerLy 1011
|||||
6318 ATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAATAA 6367
|||||
1011 sProLeuLysGlyLys..... 1016
6368 ACCTGACGCTGTACTTGTGTCATGACCCAGTGGAGAGTGGGATATCAAG 6417
|||||
1017 .....IleLeuThrPheThrGln.....SerAsp...Lys 1025
6418 AGTCTGTTGAAGGAAAGCAAAAGCAACACCACTGATGACAGTCATGA 6467
|||||
1026 GluAlaLeuLeuSerArgGly...TyrAlaAspValHisThrValHisG 1041
6468 AGCACAGGGAAACATTCAGTGTGATGTTGTTTGTAGGACGAGAAAG 6517
|||||
1041 uValGlnGlyGluThrTyrAlaAspValSerLeuValArgLeuThrProt 1058
6518 CCATGACTCCCTATTCACTAAA...CAACCGCATATACTTGTGTTG 6564
|||||
1058 hrProValSerIleIleAlaArgAspSerProHisValLeuValSerLeu 1074
6565 TCCGACACACACGCTCCTGTTTATGCGCT.....CT 6599
|||||
1075 SerArgHisThrLysSerLeuLysTyrThrValValMetAspProLe 1091
6600 GAGCTCAAAAGTTGGACGAT.....AAGTGGCGACATATATTAGCGAC 6642
|||||
1091 uValSerIleIleArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107
seq_name: SwissProt_39:RRPO_CGMVS
seq_documentation_block:
ID RRPO_CGMVS STANDARD; PRT; 1648 AA.
AC P19523; Q83208; P89877; P90356;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
OS Cucurbit green mottle mosaic virus (watermelon strain SH) (CGMV), and
OS Cucurbit green mottle mosaic virus (watermelon strain W) (CGMWV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SH;
RX MEDLINE=91311400; PubMed=1856687;
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA Sato T., Motoyoshi F., Nishiguchi M.;
RT "The complete nucleotide sequence of cucumber green mottle mosaic
RT virus (SH strain) genomic RNA.";
RN J. Gen. Virol. 72:1487-1495(1991).
RN [2]
RP REVISIONS.
RC STRAIN-SH;
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA Sato T., Motoyoshi F., Nishiguchi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1496-1648 FROM N.A.
RC STRAIN-W;
RX MEDLINE=89073773; PubMed=3201760;
RA Saito T., Imai Y., Meshi T., Okada Y.;
RT "Interviral homologies of the 30K proteins of tobamoviruses.";
RL Virology 167:653-656(1988).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA

```





```

:::||||| ||| ||| ||||| ::| |||
656 AsnLysGlyAlaLeuValGluTyrLysValGlyThrSerMetThrLeuPr 672
4219 .....CAAGAAAGCGGAAGAGCTGCAGATGGTAACACAGCACAAA 4258
|||:::|||||:::|
672 onLaThrTrpAlaGluLysGlyLysAlaVal..... 682
4259 CCGCGGTCAACGAAGCGCACAGGAGCCAGTACAGTCCAGTCTGTGTGAGT 4308
|||:::|||||:::|
683 .....LeuProLeuSerGlyLeuValArg 691
4309 TCGCCNACAGGTGATATTCCAAAGGTCACCGAGTCCGAGGTACATGCTCA 4358
:::|||||
692 LysProGln.....PheSe 696
4359 CAAGAAGCTGAACAAGAACTACCATTTGGCGACTGTTTCGGCGCGCACGC 4408
||| ::| ::| |||
696 rLysProLeuaspGlu..... 702
4409 CAATGTCGATGAGAAACCCGCCCAAGTGTACGACTCGTGTGTGAAG 4458
|||:::|
703 .....AspAspLeuArgLeuSerAsnMetAsnPheLysValSer 716
4459 ATAATTGACAAGGCAAGCGCTCGCTCATGTGGCTGAGAAAAACAGGT 4508
::| |||:::|||||:::|
717 AspLeuLysLeuLysLysThrIleThrProValValTyrThrGlyThrIl 733
4509 ACAAGTCAGCAGCCCAACACAGAGAGTTTGACGATCAATGAAGGCAAG 4558
::| ::| |||
733 eArgGluArgGlnMetLys..... 739
4559 CCGGTAAACAGCTTTGCATGTTTAGAACGTGTCTCGCGGTGTGCAGCTG 4608
739 ..... 739
4609 GATGTGTACAGAGCGACTATCGCCACAGGTTCTCAACGGCATTTAC 4658
::|
740 .....As 740
4659 CTTTTCGCATAACTTGAAGGAGGAGTGGCGTCTTTTCTCAAGCTGG 4708
:::||||| ||| ||||| ::| ||| |||
740 niyrlleaspyrLeu.....SerAlaSerLeuGlySerThrLeuG 754
4709 GTAGGGGTATACCTATATAGTGGTAGCCATGTTTCATCAGGTTGGCCT 4758
|| ||| ||| |||
754 ly.....AsnLeuGluArgIleValArgSerAspTrpAsn 765
4759 CGTCCCTAGAGGATATCTTAACGCCAATTAAAGTACCAACGCTTTCGA 4808
::| |||:::|
766 GlyThrGluGluSerMetGlnThrPheGlyLeuTyr..... 777
4809 CCACGTGTTTAGTCAGAAAGTACAAGATGGGTGGAGCGGTACCAATCCACG 4858
777 ..... 777
4859 CTGATGACGAGGAGTGTCATCCATCAGATAACCCCTATCTTGACGGTCAAT 4908
||| |||:::|
778 ..AspCysGluLysCys..... 782
4909 CTCGTGGGGAAGGCNAACCTTCTCGACTAAGTGCAGGAAGGTGTAAGGT 4958
782 ..... 782
4959 CATGTCATAAAGCTAGCTTCGGGTGACTATTTCCTTATGCTTCGGGT 5008
:::|||||:::|
783 .....LysTrpLeuLeuProAla.... 789
5009 TTCAAGGACGACACTGCATTACAGTAAACTCCATCGACGAAGGCGCATC 5058
:::| ||| ::|
790 ..GluLysLysHisAlaTrpAlaVal.....Val 798
5059 AGTTTGACGTTACGGGCACTCGGGCGCTTTTGGTGTAGCAGGATGTT 5108
::| ::| ||||| ::|
```

```

799 LeuAlaSerAspAspThrThrArgIleIlePhe..... 809
5109 GCAGTTAGCCGCGCGTCTCGATGAGAAGTCACCAAGTGTTCCAAACC 5158
:::| ||| :::| ||| ||| |||
810 .....LeuSerTyrAspGluSer...GlySerPro.... 818
5159 AGCAACCAACAGAGCAAGGTGCTACCAACAATCACACCAAAATCGGG 5208
||| ||| |||:::|
819 .....IleIleAspLysLysasn 824
5209 GGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGAC 5258
||| :::| ||| |||
825 TrpLysArgPheAlaValCysSerGluThrLysVal..... 836
5259 ATACTCGATATGTCGGAACAAGATTACCTTAGGAAGTGTGAGTGGCTCA 5308
|||||:::| ::| ||| ||| |||
837 ..TyrSerVal.....IleArgSerLeuGluValLeu. 846
5309 GGGCTGATAATCCAGTGTGCTCTTGAACCTGACTACACCCCAATGACA 5358
846 ..... 846
5359 TTTGAAGTGGTTAAACCGGACCTCTGAAGATCGCGTGGAGTACTT 5408
:::| |||:::| |||:::|
847 .....AsnLysGluAlaIleValAsp..... 853
5409 GAAGTATCTGCTATAGGCAATTGAGAGGACATACAGGCGTGTGCTTATGG 5458
853 ..... 853
5459 CTAGAAATATTCCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT 5508
853 ..... 853
5509 CAAGTTTATGAATCACTACCGGCGTTCACGTTTACAAAGTCGGGCACACA 5558
||||| |||:::|
854 .....ProGlyValHisIle..... 858
5559 TCTCATTTTTCATTCACACAAAGCGCTGGTGTGAGAGACCTACCGT 5608
::| |||:::|
859 .....ThrLeuValAspGlyVal..... 864
5609 ACGTACTCATAGTGAAAGGTATCTTTACCAGGGCAAGAGATGTCGAC 5658
864 ..... 864
5659 CCGGTGGTAGCTTTGGCGCACAACTCTGTTCTATCGGACGATATACTGTT 5708
864 ..... 864
5709 TTTCCACGATGCCATTAAATTTGATAGGTGCACTGAAAGTCGCTCGATGCG 5758
864 ..... 864
5759 GCATGGTGGCGAATCGTTTAAAGTCCTTCGAATATAAGTCTATAATGTT 5808
864 ..... 864
5809 CCCCAGGTGGCGGTAAGACGACGACGCTTAGTC..... 5841
||||| |||||:::| :::|
865 ..ProGlyCysGlyThrAlaGluIleIleAlaArgValAsnTrpLys 880
5842 .GACCAATTTCGTTAAGTCACCAATAGCACGCCACCATACGGGTAATG 5890
:::| ||| :::| |||:::|
880 sThrAspLeuValLeuThrProGlyArgGluAlaAla..... 892
5891 TGGGAAGTCTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAAT 5940
:::| |||:::| |||:::|
893 .....AlaMetIleArgArgAlaCysAla 901
5941 TTGGAAGTCTCAACAGTGTACACAGTTAACTCCAGGGTGGTAACTT 5990
||||| |||||:::| ||| |||:::|
902 LeuHisLysSerProValAlaThrAsnValArgThrPheAspSe 918
```



```

5991 TATCGTC.....AGGCAATGATATAAA.....ACGCTTTTGGTGGATG 6028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
918 rPheValMetAsnArgLysIlePheLysPheAspAlaValTyrValAsp 935
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6029 AGCTGCACATGATGACCTAGGCTTACTACACATAGCGCTCTTCGCAACC 6078
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 luGlyLeuMetValHisThrGlyLeuLeuAsnPheAlaLeuLysIleSer 951
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6079 GCGCGCTCGGAAGCCCTCTTTTGGAGACATAAATCAGATACCATTCAT 6128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
952 GlyCysLysLysAlaPheValPheGlyAspAlaLysGlnIleProPheII 968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6129 AACACGGGAGAAAGCTTTAGGATGATGCTGCTGTTTTTGTTCCAAAGA 6178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
968 eAsnArg.....ValMetAsnPheAsp.....TyrProLysG 979
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6179 AG.....GAAAGCGTGTATACACATCTTCTAAATCGTAC 6210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
979 luLeuArgThrLeuIleValAspAsnValGluArgArgTyrValThrHis 995
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6211 AGGTGTCGCTAGATGTTGCTACTTGTCTCTCAATGACCGTAAGGG 6260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
996 ArgCysProArgAspValThrSerPheLeuAsnThrIle..... 1008
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6261 AACGGAAGTGTACCTGAAAGCTGTACCGGTAAAGGACAAACACAG 6310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1009 .....TyrLysAlaAlaValAlaThrThrSerProV 1019
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6311 TACTAAGATCGCTGCCAAA.....AGG 6333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1019 alValHisSerValLysAlaIleLysValSerGlyAlaGlyIleLeuArg 1035
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6334 CCAATTGGAACCACTGATGACGTAGCTGAATAAAGCGTGACGTGACTT 6383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1036 Pro.....GluLeuThrLysIleLysGlyLysIle...II 1046
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6384 GTCCATGACCCAGCTTGAGAAAGTCGGATATGAAGAGTGTCTTGAAGGAA 6433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1046 eThrPheThrGlnSerAspLysGlnSerLeuIleLysSer..... 1059
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6434 AAGCAAAAGAACACCACTGATCAGTCAGTACAGCACAGGAGGAAACAA 6483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 .....GlyTyrAsnAspValAsnThrValHisGluIleGlnGlyGluThr 1074
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6484 TTCAGTGTGTGTATGTTAGGACGAAGAACCGCATGACTCCCTATT 6533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 PheGluGluThrAlaValAlaArgAlaThrProThrProIleGlyLeuII 1091
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6534 CACTAAA...CAACCCATATACCTTGTGTTGTCGACACACACAGCT 6580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1091 eAlaArgAspSerProHisValLeuValAlaLeuThrArgHisThrLysA 1108
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6581 CACTGCTTTATGCCGCTCTCAGCTCAAGTTCGACCATAGGTTCGCACAC 6630
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 laMetValTyrThrVal.....ValPheAspAlaValThrSer 1121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6631 TATATTAGGACGCGCTCACCTCAATCAGTATCCGACGCTTGTCTCACAC 6680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122 IleIleAlaAspVal.....GluLysValAspGlnSerIleLeuThrMe 1136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6681 GTTCGCC 6687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1136 tPheAla 1138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq\_name: SwissProt\_39:RRPO\_PMVJ

seq\_documentation\_block:

```

ID RRPO_PMVJ STANDARD; PRT; 1612 AA.
AC P89657; P90347;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)

```

```

DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Pepper mild mottle virus (strain Japan) (PMMV-J).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Kirita M., Akutsu K., Watanabe Y., Tsuda S.;
RT "Nucleotide sequence of the Japanese isolate of pepper mild mottle
   tobamovirus."
RL Ann. Phytopathol. Soc. Jpn. 0:0-0(1997).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
   RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
   CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
   BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000709; BAA19167.1; -
DR EMBL; AB000709; BAA19166.1; -
DR INTERPRO; IPR000606; -
DR INTERPRO; IPR001788; -
DR INTERPRO; IPR002588; -
DR PFAM; PF00978; RNA_dep_RNAPol2; 1.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Transference; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1117 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 835 842 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183303 MW; E7E66CEBED81FE4C CRC64;

```

```

alignment_scores:
  Quality: 297.00      Length: 854
  Ratio: 0.754        Gaps: 40
  Percent Similarity: 46.136      Percent Identity: 22.834

```

alignment\_block:

US-09-301-906-3 x RRPO\_PMVJ

Align seg 1/1 to: RRPO\_PMVJ from: 1 to: 1612

```

4522 CCCAAACAGAGGAGTTTGACGATC..... 4545
      |||||:|||||:|||||:|||||:
366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuG1 382
4546 .AATGAAGGCAAG.....GCCGGTAAACAGCTTTCATGTTAGAACGT 4588
      |||||:|||||: ||| |||:|||||: ||| |||:
382 nAsnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAsp 399
4589 GTTCCTCGGTGTGCACGTGGATGTGTACACGACGACTATGCCACC 4638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu 415
4639 AGGTTCTCAACGCATTTACCTTTGTCGATACTTTGAAGGGAGG..... 4683
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValII 432
4683 ..... 4683
432 eIleAsnGlyValThrAlaArgSerGluTyrAspValAspLysAlaLeuL 449
4684 .....AGTGGGCTTTTTC.....TCAAAGCTGGGT..... 4710
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 euGlnSerLeuSerMetThrPhePheLeuGlnThrLysLeuAlaMetLeu 465

```

```
4711 .....GAGGGTATACCTATAATGGTGTAGCCATGT 4742
      : : : : : : : : : : : : : : : : : : : :
466 LysAspSerLeuValValGlnLysPheGlnValHisSerLysSerLeuTh 482
      : : : : : : : : : : : : : : : : : : : :
4743 TTCATCAGGGTGG.....CCTC 4759
      : : : : : : : : : : : : : : : : : : : :
482 rglutyrValtrpAspGluIleThrAlaAlaPheHisAsnCysPheProt 499
      : : : : : : : : : : : : : : : : : : : :
4760 GTGCCCTAGAGGATATCTTAAGC..... 4782
      : : : : : : : : : : : : : : : : : : : :
499 hrIleGlyGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys 515
      : : : : : : : : : : : : : : : : : : : :
4783 .....GCAATTAAAGTACCCAGCGTC.....TTCGACCACCTGTT 4817
      : : : : : : : : : : : : : : : : : : : :
516 AlaLeuGluIleLysValProAspLeuIleThrPheHisAspArgLe 532
      : : : : : : : : : : : : : : : : : : : :
4818 AGTCAGAAAGTACAAAGATGGGTGGAGCGGTACCA.....TTCACG 4858
      : : : : : : : : : : : : : : : : : : : :
532 uValLysGluIleLysSerValGluMetProValLeuAspValLysL 549
      : : : : : : : : : : : : : : : : : : : :
4859 CTGATCAGCAGAGTCCATCCATCAGATAACCCCTATCTTGACGGTCAAT 4908
      : : : : : : : : : : : : : : : : : : : :
549 ySerLeuGluGluAlaGluValMetTyrAsnAlaLeuSerGluIleSer 565
      : : : : : : : : : : : : : : : : : : : :
4909 CTCGTGGGAGGCAAAAC.....TTCTCGACTAAGTG 4940
      : : : : : : : : : : : : : : : : : : : :
566 IleLeuLysAspSerAspLysPheAspValAspValPheSerArgMetCy 582
      : : : : : : : : : : : : : : : : : : : :
4941 CAGGAAGGGTGGT.....AAGGTGATGGTCATAA 4969
      : : : : : : : : : : : : : : : : : : : :
582 sAsnThrLeuGlyValAspProLeuValAlaAlaLysValMetValAlaV 599
      : : : : : : : : : : : : : : : : : : : :
4970 AGGTAGCTTCGGGTGACTATTTCTTATGCCTTGGCTTTCAAAGGAGC 5019
      : : : : : : : : : : : : : : : : : : : :
599 alValSer..... 601
      : : : : : : : : : : : : : : : : : : : :
5020 CACTTGGCATTCAGTAACTCCATCAGCAAGGGCGCATCAGTTTGACGTT 5069
      : : : : : : : : : : : : : : : : : : : :
602 .....AsnGluSerGlyLeuThrLeuThrPh 610
      : : : : : : : : : : : : : : : : : : : :
5070 CAGGGCACTCGGCGCTCTTTGGTGTAGGAGGATGTTGCAAGTTAGCG 5119
      : : : : : : : : : : : : : : : : : : : :
610 eGluArgProThrGluAla...AsnValAlaLeuAlaLeuGln..... 623
      : : : : : : : : : : : : : : : : : : : :
5120 GCGGCGTGTCCGATCAGAGTCACCAAGTGTTCCAAACCAGCAACCACAG 5169
      : : : : : : : : : : : : : : : : : : : :
624 .....ProThrIleThrSerLys 629
      : : : : : : : : : : : : : : : : : : : :
5170 AGCCAAAGTGTACCAGAACATACACAAATCGGGGGCAGGCTCT 5219
      : : : : : : : : : : : : : : : : : : : :
630 GluGluGlySerLeuLysIleValSerSerAspValGly..... 642
      : : : : : : : : : : : : : : : : : : : :
5220 ATCTGAGGAAGTGTAGGAAGTCAAGGGAGGTGCGACATCTCGATA. 5268
      : : : : : : : : : : : : : : : : : : : :
643 .....GluSerIleLeuGluValValArgLysSerGluIleSerMetL 658
      : : : : : : : : : : : : : : : : : : : :
5269 .....TGGTCCGAACAAGATTACGTTAGGAAGTCTGAG 5301
      : : : : : : : : : : : : : : : : : : : :
658 euGlyLeuThrGlyAsnThrValSerAspGluPheGlnArgSerThrGlu 674
      : : : : : : : : : : : : : : : : : : : :
5302 TGGCTCAGGGGTGATAATCCAGTGTGCTCTTGAACCTGACTACACCCC 5351
      : : : : : : : : : : : : : : : : : : : :
675 .....IleGluSerLeuGln..... 679
      : : : : : : : : : : : : : : : : : : : :
5352 ATGACATTGAAGTGGTTAAACCGGAGCC.....TCTGAAGATG 5392
      : : : : : : : : : : : : : : : : : : : :
680 .....GlnPheHisMetValSerThrGluThrIleIleArgLysGlnMetH 695
      : : : : : : : : : : : : : : : : : : : :
5393 CCGTCGTGCACTTGAAGTATCTGGCTATAGGCATTTGACAGGACATAC 5442
      : : : : : : : : : : : : : : : : : : : :
695 isAlaMetValIleThrGlyProLeuLysValGlnGlnCysLysAsnTyr 711
```

```
980 GlnLysTyrGluGlyGlnValMetCysThrSerValThrArgSerVa 996
    |||  ...  |||  ...
6276 CCTGAAAGGTGCTAGCGGTAGGCAACCAACAGTAGTAGTCGCTGT 6325
    |||  ...  |||  ...
996 lSerHisGluValIleGlnGly.....AlaAlaValMetAsnProVal 1011
6326 CCAAAAGGCCAATTCGAACACCTGATGACCTAGCTGAAATAACGCTGAC 6375
    |||  ...  |||  ...
1011 erLysProLeuLysGlyLysVal..... 1018
6376 GTGTACTTGTGCATGACCCAGTTCGAGAGTCGATATGAAGGTCGTT 6425
    |||  ...  |||  ...
1019 .....lIleThrPheThrGlnSerAspLysSer.....LeuLe 1029
6426 GAAGGAAAGGAAAGAAACACCACTGATGACAGTGCATGAAGCACAGG 6475
    |||  ...  |||  ...
1029 uLeuSerArgGlyTyrGlu...AspValHisThrValHisGluValGln 1045
6476 GAAACATTCACGTGCTGCTATTCTTTAGGACGAAGAAGCCGATGAC 6525
    |||  ...  |||  ...
1045 lYgluThrPheGluAspValSerLeuValArgLeuThrProThrProVal 1061
6526 TCCTATTCTACTAAACAA...CCGCATATACTTGTGGTGTTCGAGACA 6572
    |||  ...  |||  ...
1062 GlyIleIleSerLysGlnSerProHisLeuValSerLeuSerArgHl 1078
6573 CACAGCTCCTACTGTTTAT.....GCCGCTCTGAGTCA 6607
    |||  ...  |||  ...
1078 sThrArgSerIleLysTyrThrValValLeuAspAlaValValServ 1095
6608 ACTTGCACGATAG.....GTCGGCACATATATTAGCGACGCTCACCT 6651
    |||  ...  |||  ...
1095 alLeuArgAspLeuGluCysValSerSerTyrLeuLeuAspMetTyrLys 1111
6652 CAATCAGTATCC 6663
    |||  ...  |||  ...
1112 valAspValSer 1115
```

seq\_name: SwissProt\_39:RRPO\_PMVVS

```
seq_documentation_block:
ID  RRPO_PMVVS STANDARD; PRT: 1612 AA.
AC  P29098; 084924;
DT  01-DEC-1992 (Rel. 24, Created)
DI  01-OCT-2000 (Rel. 40, Last sequence update)
DE  01-OCT-2000 (Rel. 40, Last annotation update)
DE  RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
OS  METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OC  Pepper mild mottle virus (strain Spain) (PMV-S).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92113528; PubMed=1765765;
RA  Alonso E., Garcia-Luque I., de la Cruz A., Wicke B.,
RA  Avila-Rincon M.J., Serra M.T., Castresana C., Diaz-Ruiz J.R.;
RT  "Nucleotide sequence of the genomic RNA of pepper mild mottle virus,
    a resistance-breaking tobamovirus in pepper.";
RL  J. Gen. Virol. 72:2875-2884(1991).
CC  -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
    RNA REPLICATION
CC  -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
    CAPPING AND AN RNA HELICASE.
CC  -!- MISCELLANEOUS: REATHROUGH OF THE TERMINATOR CODON UAG OCCURS
    BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
```

```
CC  -----
DR  EMBL; M81413; AAB02334.1; -.
DR  EMBL; M81413; AAB02335.1; -.
DR  PIR; JQ1312; WMTMPV.
DR  INTERPRO; IPR00606; -.
DR  INTERPRO; IPR001788; -.
DR  INTERPRO; IPR002588; -.
DR  PFAM; PF00978; RNA_dep_RNAPol2; 1.
DR  PFAM; PF01443; Viral_helicase1; 1.
DR  PFAM; PF01660; Vmethyltransf; 1.
KW  transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT  CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT  CHAIN 1 1117 METHYLTRANSFERASE/RNA HELICASE.
FT  NP_BIND 835 842 ATP (POTENTIAL).
SQ  SEQUENCE 1612 AA; 183265 MW; 465677F4924802BF CRC64;
```

```
alignment_scores:
    Quality: 297.00      Length: 854
    Ratio: 0.754        Gaps: 40
    Percent similarity: 46.136      Percent Identity: 22.834

alignment_block:
US-09-301-906-3 x RRPO_PMVVS
Align seg 1/1 to: RRPO_PMVVS from: 1 to: 1612
```

```
4522 CCCAAACAGAGAGTTTGACGATC..... 4545
    |||  ...  |||  ...
366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuGl 382
4546 .AATGAAGCAAG.....CCCGTAAACAGCTTTGCATGTTAGAACGT 4588
    |||  ...  |||  ...
382 nAsnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAsp 399
4589 GTTCCTCGCGTGTGCAGCTGATGTGTACACGACGACTATCGCCACC 4638
    |||  ...  |||  ...
399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu 415
4639 AGGTTCTCAAACGCTTTACCTTTGTCGATACTTGAAGGGAGG..... 4683
    |||  ...  |||  ...
416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValIl 432
4683 ..... 4683
432 eIleAsnGlyValThrAlaArgSerGluTrpAspValAspLysAlaLeuL 449
4684 .....ACTGCGGTCTTTTC.....TCAAAGCTGGGT..... 4710
    |||  ...  |||  ...
449 euGlnSerLeuSerMetThrPheLeuGlnThrLysLeuAlaMetLeu 465
4711 .....GAGGGGTATACCTATAATGGTGGTAGCCATGT 4742
    |||  ...  |||  ...
466 LysAspAspLeuValValGlnLysPheGlnValHisSerLysSerLeuTh 482
4743 TTCATCAGGGTGG.....CCTC 4759
    |||  ...  |||  ...
482 rGluTyrValTrpAspGluIleThrAlaAlaPheHisAsnCysPheProt 499
4760 GTCCCTCAGGATATCTTAACG..... 4782
    |||  ...  |||  ...
499 hrIleLysGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys 515
4783 .....GCATTAAAGTACCAAGCGTC.....TTCGACCACCTGTTT 4817
    |||  ...  |||  ...
516 AlaLeuGluIleLysValProAspLeuTyrValThrPheHisAspArgLe 532
4818 AGTCGAGAAGTACAAGATGGGTGGAGCGGTACCA.....TTCACG 4858
    |||  ...  |||  ...
532 uValLysGluTyrLysSerValGluMetProValLeuAspValLysL 549
4859 CTGATCAGCAGGAGTCTCTCCATCAGATAACCCCTATCTTCGACGGTCAAT 4908
    |||  ...  |||  ...
```



```
6526 TCCCTATTCTAAACAA...CCGCATATATCTTGTGTTGTGAGACA 6572
.....:|||||:|||||:|||||:|||||:|||||:|||||:
1062 GlyIleSerLysGlnSerProHisLeuValSerLeuSerArgH1 1078
|||||:|||||:|||||:|||||:|||||:|||||:
6573 CACAGCTCACTGTTTAT.....GCCGCTCTGAGCTCAA 6607
|||||:|||||:|||||:|||||:|||||:|||||:
1078 sThrArgSerIleLysTyrThrValValLeuAspAlaValServ 1095
|||||:|||||:|||||:|||||:|||||:|||||:
6608 AGTTGGACGATAG.....GTGGCACAATATATTACGAGCGTCACT 6651
|||||:|||||:|||||:|||||:|||||:|||||:
1095 alLeuArgAspLeuGluCysValSerSerTyrLeuLeuAspMetTyrLys 1111
|||||:|||||:|||||:|||||:|||||:|||||:
6652 CAATCACTATCC 6663
:|||||:
1112 ValAspValSer 1115
:|||||:
```

seq\_name: SwissProt\_39:VIA\_BBMV

seq\_documentation\_block:

ID VIA\_BBMV STANDARD; PRT; 966 AA.

AC Q00020;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE 1A PROTEIN [INCLUDES: HELICASE; METHYLTRANSFERASE].

OS broad bean mottle virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;

OC Bromovirus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BADEN;

RX MEDLINE-92074218; PubMed-1962437;

RA Dzanott A.M., Bujarski J.J.;

RT "The nucleotide sequence and genome organization of the RNA-1 segment in two bromoviruses: broad bean mottle virus and cowpea chlorotic mottle virus.";

RL Virology 185:553-562(1991).

CC 1- FUNCTION: MAY BE INVOLVED IN THE REPLICATION OF THE VIRUS.

CC CONTAINS AN HELICASE DOMAIN AND A METHYLTRANSFERASE DOMAIN. THE

CC METHYLTRANSFERASE DOMAIN IS PROBABLY INVOLVED IN VIRAL RNA

CC CAPPING.

CC 1- SIMILARITY: TO 1A PROTEIN FROM CCMV, CMV, PSV AND TAV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M65138; AAA42740.1; -

DR PIR; A41699; P1BVBB.

DR INTERPRO; IPR000606; -

DR INTERPRO; IPR002588; -

DR PFAM; PF01443; Viral\_helicase1; 1.

DR PFAM; PF01660; Vmethyltransf; 1.

KW Helicase; ATP-binding; Transferase; Methyltransferase.

FT NP\_BIND 690 697 ATP (POTENTIAL).

SQ SEQUENCE 966 AA; 109621 MW; DF592681D7231C8D CRC64;

alignment\_scores:

Quality: 290.00 Length: 775

Ratio: 0.773 Gaps: 33

Percent Similarity: 48.387 Percent Identity: 23.613

alignment\_block:

US-09-301-906-3 x VIA\_BBMV ..

Align seg 1/1 to: VIA\_BBMV from: 1 to: 966

4580 TTAGAAGCTGTTCTCGGTGTCACCTCGATG.....TGATAC 4617

```
|||||:|||||:|||||:|||||:|||||:|||||:
275 LeuGluArgMetLeuLeuLysCysSerIleMetThrTyrLysIleIleAl 291
:||||:||||:||||:||||:||||:||||:||||:
4618 AACGAAGCGACTATCGCCACCGAGTCTTCAACGCATTT ACCTTTGTGCG 4666
:||||:||||:||||:||||:||||:||||:||||:
291 aThrAsnValArgCysProGluSerLeuArgHisCysIleTrpPheG 308
|||||:|||||:|||||:|||||:|||||:|||||:
4667 ATAACCTTGAAGGAGGAGTGGCTCTTTTCTCAACGCTGGTGGGAGGG 4716
:||||:||||:||||:||||:||||:||||:||||:
308 luAsnIleSerGlnTyrLeuAlaVal.....GlnIleProIleGly 321
|||||:|||||:|||||:|||||:|||||:|||||:
4717 TATACCTATAATGCTGCTAGCCATGTTTCATCAGGTGGCGCTCGTGCCT 4766
|||||:|||||:|||||:|||||:|||||:|||||:
322 TyrAsnLeuAsn.....AspTrpLysThrValAr 331
|||||:|||||:|||||:|||||:|||||:|||||:
4767 AGAGGATATCTTAACGGCAATTAAGTACCACCAAGCGTCTTCGACCACCTGT 4816
:||||:||||:||||:||||:||||:||||:||||:
331 gValAlaArgAlaThrValArgGluValGluIleSerPheArgCysP 348
|||||:|||||:|||||:|||||:|||||:|||||:
4817 TAGTGCACAAGTACAAGATGGGTGGAGCGGTACCATTCCAGCTGATGAC 4866
:||||:||||:||||:||||:||||:||||:||||:
348 heLysGlu.AsnLysAspTrpThrGluAsnMetArgSerValAlaSerI 364
|||||:|||||:|||||:|||||:|||||:|||||:
4867 GAGGACTCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGG 4916
:||||:||||:||||:||||:||||:||||:||||:
364 eLeuSerAlaLysSerSerThrValIle.IleAsnGlyGlnSerIleMet 380
|||||:|||||:|||||:|||||:|||||:|||||:
4917 GAAGGCAAACTTCGACTAAGTGCAGGAAGGTGTTAAGTGCATGCTGCTCA 4966
:||||:||||:||||:||||:||||:||||:||||:
381 SerGlyGluArgLeuAspValLeuGlu..... 389
|||||:|||||:|||||:|||||:|||||:|||||:
4967 TAAACGTAGCTTCGGGTGACTATTTCTTATGCTTCGGTTCCTTCAAGG 5016
:||||:||||:||||:||||:||||:||||:||||:
390 .....TyrHisLeuValAlaPheSerLeut 398
|||||:|||||:|||||:|||||:|||||:|||||:
5017 ACGCACTTGCTTCACTAACTCCATCGCAGGAAGGCGCATCAGTTTCAC 5066
:||||:||||:||||:||||:||||:||||:||||:
398 hrLeuAsnLeuTyrGlnLysTyrGluLysLeuArgAsnPheGlnGlyGlu 414
|||||:|||||:|||||:|||||:|||||:|||||:
5067 GTTCAGGCGCAAC.....TCGGCGCGTCTT...TGGTG 5095
:||||:||||:||||:||||:||||:||||:||||:
415 LeuGluTrpLysGlyTyrPalaAsnHisPheLysThrArgLeuTrpTrpCy 431
|||||:|||||:|||||:|||||:|||||:|||||:
5096 TAGCAGGATGTGCTACCTTACCGCGCGC..... 5124
:||||:||||:||||:||||:||||:||||:||||:
431 s.GlyArgThrValSerThrGluGlyGlyPheLeuArgAsnPheLeuAla 447
|||||:|||||:|||||:|||||:|||||:|||||:
5124 ..... 5124
|||||:|||||:|||||:|||||:|||||:|||||:
448 AspLysIleProTrpLeuLysLeuAsnThrTyrAlaAspSerLeuAspPh 464
|||||:|||||:|||||:|||||:|||||:|||||:
5125 .....GTGTGGGATGAGAAGTCA.....CCAGGTGTTCCAA 5155
:||||:||||:||||:||||:||||:||||:||||:
464 elleThrLysIleSerGluValGluSerPheGluValAspSerValPro 481
|||||:|||||:|||||:|||||:|||||:|||||:
5156 ACCAGCAACCAACAGAGC...CAAGGTGCTACCAGAACATCACCACAAA 5202
:||||:||||:||||:||||:||||:||||:||||:
481 hrSerArgLeuArgSerPhePheGlnLysGluGluAsnIleValGluArg 497
|||||:|||||:|||||:|||||:|||||:|||||:
5203 TCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGCAG 5252
:||||:||||:||||:||||:||||:||||:||||:
498 AlaAlaSerGluIleMetSerAlaAsnAlaArgIleAlaLys..... 512
|||||:|||||:|||||:|||||:|||||:|||||:
5253 GTCGACATACATCGATGTTGTCGGAACAAGATTACGTTAGGAAGTGTGAGT 5302
:||||:||||:||||:||||:||||:||||:||||:
513 .....LysAlaGluM 516
|||||:|||||:|||||:|||||:|||||:|||||:
5303 GGCTCAGGGCTGATAATCCAGTGTGCTGCTTGAACCTGACTACACCCCA 5352
:||||:||||:||||:||||:||||:||||:||||:
516 etSerLysGluPheAspPheValAspAlaProGluGluPheAlaPro 532
|||||:|||||:|||||:|||||:|||||:|||||:
5353 .....ATGACATTTGAAGTGGTTAAACCGGGGACCTCTCAAGATGCCCT 5396
:||||:||||:||||:||||:||||:||||:||||:
```

533 GluAspValValGluValIleAsnThrProValThrGlnAspValLy 549  
5397 CGTGGAGTACTTTAAAGTATCTCGGTATAGGCATTACAGAGACATACACAGG 5446  
::: :::  
549 sLeuArgInSerLys.....ProGluThrAlaArgS 560  
5447 CQTTCCTTAATG.....GCTAGAATAATTCGCCGTAC.... 5477  
::: :::  
560 erileValIleuAspProAspAlaValLeuLysAsnGlyAlaIleAsnGlu 576  
5478 .....TACC GCCGAAGTGTTCTGAAGTAGTACCTAATCAACAGTTTATG 5518  
::: :::  
577 pheAlaAspTyrSerLysArgLeuHisGluAsnThrValSerAsnLeu.A 593  
5519 AATCACA.....CCGGCTTTCACGTTTACCAAGTCG 5550  
::: :::  
593 rgHisLeuTrpThrLeuMetGlyCysArgGlyAsnGluIleHisAsnLys 609  
5551 GCACAGATCATTTTTTCATTCACACAGAACGGCTTGGCTGGAGAGA... 5598  
::: :::  
610 SerValAlaGluThrTyrHisArgValAspAspMetValAsnValHisPh 626  
5599 .....GACCTACCGTAGCTACTCATAGCTGAAA 5626  
::: :::  
626 eProAsnGlyHisTrpMetTyrProLeuLysTyrGluTyrThrValGlyT 643  
5627 AAGGTATCTTTACCAAGGCCAAAGATGTCGACGGGTGGTAGCTTGGGC 5676  
::: :::  
643 yrAsnAspGlyClyLeuGlyLysPheGluAsnGluLeuTyrValVal 659  
5677 GCAAATCTCTCGTATGCGCACATATACTGGTTTCCACCATGCCATTAA 5726  
::: :::  
660 AspLysThrCysSerCysAlaAsnAlaLysAlaIleAlaAspAlaCysLy 676  
5727 TTTGATAGTGCACCTGAAGTCGCTCGATCGGCATCGTGGCGCAATCGT 5776  
::: :::  
676 slsValSerAla.....ProThrCysSerValVal..... 686  
5777 TTAAGTCCTTCGAATATAAGTGCTAATAGTCCCCAGGTGGCGGTAAG 5826  
::: :::  
687 .....MetValAspGlyValAlaGlyCysGlyLys 696  
5827 ACACAGCAGCTTAGTGACGAATTCGTTAAGTCACCCCATAGCACAGCCAC 5876  
::: :::  
697 ThrThrAlaIleLysGluThrPhe.....ArgPheGluLysAspIleIl 711  
5877 CATTAAGCTAATGTGGGAAGTTCTGAGGACATAAATATGCGCGTGAAGA 5926  
::: :::  
711 eValThrAlaAsnArgLysSerAlaGluAspValArgLysAlaIle.... 726  
5927 AGAGAGATCCGAATTTGGAAGGTCTCAACAGTGTACACAGTTTAACCTC 5976  
::: :::  
727 .....PheGlyAspAlaSerAspSerGluValAlaLeu 737  
5977 AGGGTGGTA.....AACTTTATCTCAGGCGAATGTATAAA... 6012  
::: :::  
738 LysValValArgThrAlaAspSerAlaIleMethisGlyLeuProGluCy 754  
6013 ....AGGTTTTGGTGATGAGGTGCACATGCATCCATCAGGCTTACTAC 6058  
::: :::  
754 shisArgLeuValAspGluAlaGlyLeuLeuHisTyrGlyGlnLeuL 7711  
6059 AAATAGGCGTCTTCGCAACCGCGCGTGGGAAGCCCTCTTTTGGAGAC 6108  
::: :::  
771 euAlaValAlaAspLeuCystLysCysSerGluValLeuAlaPheGlyasp 787  
6109 ATAAATCAGATACCAATTCATAAACCAGGAGAAGGTGTTTAGGTGGAT.. 6156  
::: :::  
788 ThrGluGlnIleSerPheLysSerArgAspAlaThrPheargMetLysTy 804  
6157 .TGTCCTGTTTTGTTCCAAGAGAGGAAGCGTTGTATACACTTCTAAAT 6205  
::: :::  
804 rCysAsnIleGluTyrAspLysArgAspIleVal.....SerLysT 818

```

6206 CCTACAGGTGTCGGTTAGATGTTTGCCTACTTGTTGTCCTCAATGACCCTGA 6255
      hrPheArgCysProGlnAspVal.....ValSerAlaValLysIle 831
      ::::||||| ||||| ::|||::: ::
6256 AGGGAAACGGAAGAAGTG.....TACCCTGAAAAGGT 6287
      ::::||||| ||| |::|
832 Leu....LysArgLysCysAlaAsnArgSerSerLysTyrrAsnGlyTrpVa 847
      :|||::| ||| ::||| ||||| |
6288 CGTTAGCGGTAGAGACAACAGTAGTAAGTACGCTGCCAAAAAGGCCAA 6337
      :|||::| ||| ::||| ||||| |
847 lserSerSerLys.....ValGluLysSerLeuSerLysSerArgl 861
      :|||::| ||| ::||| ||||| |
6338 TTGGAAACCACTGATGACCTACTGCTGAATAAAGCTGACGTGTACTTGTGC 6387
      || ::::::|::|::: ::| |||||
861 leValSerIleAsnGlnValSerMetGluLysHisLysPheTyrLeuThr 877
      :|||::| ||| ::||| ||||| |
6388 ATGACCCAGTTGGAGAACTCGGATATGAAGAGTTCGTTGAAGGAAAAAG 6437
      :|||::| ||| ::||| ||||| |
878 MetThrGluAlaAspLysAlaAlaLeuCysSerArgAlaLysAspValGl 894
      :|||::| ||| ::||| ||||| |
6438 A.....AAAGAACACCAAGTGCATGATGATGAAGCAC 6472
      :|||::| ||| ::||| ||||| |
894 yLeuAspLysThrTrpValGluSerAsnMetGluThrValHisGluAlaG 911
      :|||::| ||| ::||| ||||| |
6473 AGGGAARACATTCAGTCATGCTGTATTGTTAGGACGAAGAAGCCGAT 6522
      |||||::: ::| ||||| ||| ||:::
911 InGlyLysAlaValAspHisValValLeuValArgLeuLysSerThrLys 927
      :|||::| ||| ::||| ||||| |
6523 GACTCCCATTACACTAAACAACCGCATATACTTGTGTTGTCGAGACA 6572
      :|||::| ||| ::||| ||||| |
928 CysAspLeuPheLysSerGluGluTyrCysLeuValAlaLeuThrArgHl 944
      :|||::| ||| ::||| ||||| |
6573 CACACGCTCACTGTTTTACCGCTCTGAGCTCAAAGTTCGACGATAAG 6622
      :|||::| ||| ::||| ||||| |
944 slYsArgThrPheGluTyrLeuTyrAsnGlyAspLeuGlyGlyAspLeuI 961
      :|||::| ||| ::||| ||||| |
6623 TCGGCACATATATTAGC 6639
      :|||::| ||| ::||| ||||| |
961 leSerPheTyrValThr 966
      :|||::| ||| ::||| ||||| |

seq_name: SwissProt_39:RRPO_TMVRa
seq_documentation_block:
ID RRPO_TMVRa STANDARD; PRt; 1616 AA.
AC Q98745; O41340;
DT 01-OCT-2000 (Rel. 40, Created)
DF 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobacco mosaic virus (strain Rakkyo) (TMV-R).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:96265021; PubMed-8678834;
RA Chen J., Watanabe Y., Sako N., Ohshima K., Okada Y.;
RT "Complete nucleotide sequence and synthesis of infectious in vitro
RT transcripts from a full-length cDNA clone of a rakkyo strain of
RT tobacco mosaic virus.";
RL Arch. Virol. 141:885-900(1996).
CC CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstati
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no co
CC modified and this statement is not removed. Usage by and for commer

```

```

180  isAsnThrPheGlnThrCysGluHisGlnProMetGlnGlnSerGlyLys 196
      ...      |||...|||...      :|...:
1756  GCCGATCGGGTTTTATCGTGGATGTGTACCATATATCCCGCAGCAGGT 1805
      :|...: |||...      :|...: |||...: |||...:
197  valTyrAlaIleAlaLeuHisGlyPheTyrAspIleProAlaAspGluPh 213
      :|...: |||...      :|...: |||...: |||...:
1806  AGCAGAGGCTATGGATAAGAGAGGTGCGCTGGTTTTCCACATAGCTCTTA 1855
      :|...: |||...      :|...: |||...: |||...:
213  eGlyAlaAlaLeuLeuArgLysAsnValHisThrCysTyrAlaAlaPheH 230
      :|...: |||...      :|...: |||...: |||...:
1856  TGTTCGCCGTGGAGTTGTGTACGGTACAGGTGAAGTTTACTTGGAGAA 1905
      |||...      :|...: |||...: |||...: |||...:
230  isPheSerGluAsnLeuLeuLeuGluAspSerTyrValAsnLeuAspGlu 246
      :|...: |||...      :|...: |||...: |||...:
1906  CTCGATAGCTGGTCAACAGAGGAGGTGATACCTGCGCTACATGTGTGG 1955
      :|...: |||...      :|...: |||...: |||...:
247  IleAsnAlaCysPheSerArgAspGlyAspLysLeuThrPheSerPheAl 263
      :|...: |||...      :|...: |||...: |||...:
1956  TCAGTGTGGTGAAGT...TATGAACATTCCTCTCTAACGTAAAGCGGGT 2002
      :|...: |||...      :|...: |||...: |||...:
263  aSerGluSerThrLeuAsnTyrCysHisSerPheSerAsnIleLeuLysT 280
      :|...: |||...      :|...: |||...: |||...:
2003  TTTTCACCTTTCTTATGTACGACTTCGTCCGGGAAGCTTTTAAAGCTA 2052
      :|...: |||...      :|...: |||...: |||...:
280  yValCysLysThrTyrPheProAlaSerAsnArgGluValTyrMetLys 296
      :|...: |||...      :|...: |||...: |||...:
2053  GAGTATGAGGGATACCGTTGTGTGTACCATCATCTCACTATGTGTAGGC 2102
      |||...: |||...: |||...: |||...:
297  GluPhe..... 298
      :|...: |||...      :|...: |||...: |||...:
2103  TCAGAAGTCACTCGAACTGAGGTACGTATAGCTGGTTGGTCCGTCGT 2152
      :|...: |||...      :|...: |||...: |||...:
298  ..... 298
      :|...: |||...      :|...: |||...: |||...:
2153  TCGTGGGAAATCGCTGGTGTGTATACCTGTGTAGCTGGTTCTAGTGTG 2202
      :|...: |||...      :|...: |||...: |||...:
298  ..... 298
      :|...: |||...      :|...: |||...: |||...:
2203  TCCTTTAAGACAATAGTCCTCGATTCGGACTTTCTGCACAGGATCTATT 2252
      :|...: |||...      :|...: |||...: |||...:
299  .....LeuValThrArgValAsnTh 305
      :|...: |||...      :|...: |||...: |||...:
2253  CTAC.....GCGCTCAACACTATAGGCACATTCGAGAATAGAACGTTG 2296
      :|...: |||...      :|...: |||...: |||...:
305  rTrpPheCysLysPheSerArgIleAspThrPhe..... 316
      :|...: |||...      :|...: |||...: |||...:
2297  AGTATGCCGTTGGCGCGTTCAGGTGCGCAAAAGACCCATGTCATTACAGG 2346
      :|...: |||...      :|...: |||...: |||...:
317  .....LeuLeuTyr 319
      :|...: |||...      :|...: |||...: |||...:
2347  AGTCGCGTTGTCCACAGCAAGTTGCATATTTCTCTGATGATATGTCGGG 2396
      :|...: |||...      :|...: |||...: |||...:
320  LysGlyValAlaIleHis..... 324
      :|...: |||...      :|...: |||...: |||...:
2397  TTATTGTTGCGTGTATTGGTTCAGCGCGATTAAAGGATAGCGCGAAGAGTA 2446
      :|...: |||...      :|...: |||...: |||...:
325  .....LysSerV 327
      :|...: |||...      :|...: |||...: |||...:
2447  TTCGCTCCTATAACTTTATAAAGCCAGTGAAGGAGTCTCGCGCGGGGT 2496
      :|...: |||...      :|...: |||...: |||...:
327  aLaspSerGluGlnPheTyrThrAlaMetGluAspAlaTrpHis..... 341
      :|...: |||...      :|...: |||...: |||...:
2497  TTCAAGCTCTCTTTCACACCGTAGCGCATGTTTTCGAACACGAGTCTC 2546
      :|...: |||...      :|...: |||...: |||...:
342  .....TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355
      :|...: |||...      :|...: |||...: |||...:
2547  CGCTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
      :|...: |||...      :|...: |||...: |||...:
355  uLeuGluAspSerSerValAsnTyrTrpPheProLysMetArgAspM 372
      :|...: |||...      :|...: |||...: |||...:
2597  TTATCTCTATGCCACAGCGGTTTCATCCGTAAAGTACCTGGGTCGTGTGTT 2646
      :|...: |||...      :|...: |||...: |||...:

```

```

180  isAsnThrPheGlnThrCysGluHisGlnProMetGlnGlnSerGlyLys 196
      ...      |||...|||...      :|...:
1756  GCCGATGGGGTTTTATCGTGGATGTGTACCATATATCCCGCAGCAGGT 1805
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
197  valTyrAlaIleAlaLeuHisGlyPheIleTyrAspIleProAlaAspGluPh 213
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
1806  AGCAGAGGCTATGGATAAGAGAGGTGCGCTGGTTTTCCACATAGCTCTTA 1855
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
213  eGlyAlaAlaLeuLeuArgLysAsnValHisThrCysTyrAlaAlaPheH 230
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
1856  TGTTCGCCGTGGAGTTGTGTACGGTACGGTGAAGTTTACTTGGAGAA 1905
      |||...|||...|||...|||...|||...|||...|||...|||
230  isPheSerGluAsnLeuLeuLeuGluAspSerTyrValAsnLeuAspGlu 246
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
1906  CTCGATAGCTGGTCAACAGAGGAGGTGATACCTGCCCTACATGTGTGG 1955
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
247  IleAsnAlaCysPheSerArgAspGlyAspLysLeuThrPheSerPheAl 263
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
1956  TCAGTGTGGTGAAGT...TATGAACATTCCTCTCTAACGTAAAGCGGGT 2002
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
263  aSerGluSerThrLeuAsnTyrCysHisSerPheSerAsnIleLeuLysT 280
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2003  TTTTCACCTTTCTTATGTACGACTTCGTCCGGGAAGCTTTTAAAGCTA 2052
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
280  yValCysLysThrTyrPheProAlaSerAsnArgLysValTyrMetLys 296
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2053  GAGTATGAGGGATACCGTTGTGTACCATCATCTCACTATGTGTAGGGC 2102
      |||...|||...|||...|||...|||...|||...|||...|||
297  GluPhe..... 298
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2103  TCAGAAGTCACTCGAACTGAGGTACGTATAGCTGGTTGGTCCCGTCGT 2152
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
298  ..... 298
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2153  TCGTGGGAAATCGCTGGTGTTCATACCTGTGTAGCTGGTTCTAGTGTG 2202
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
298  ..... 298
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2203  TCCTTTAAGACAATAGTCCTCGATTCGGACTTTCTGCACAGGATCTATT 2252
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
299  .....LeuValThrArgValAsnTh 305
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2253  CTAC.....GCGCTCAACACTATAGGCACATTCGAGAATAGAACGTTG 2296
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
305  rTrpPheCysLysPheSerArgIleAspThrPhe..... 316
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2297  AGTATGCCGTTGGCGCGTTCAGGTGCGCAAAAGACCCATGTCATTACAGG 2346
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
317  .....LeuLeuTyr 319
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2347  AGTCGCGTTGTCCACAGCAAGTTGCATATTTCTCTGATGATATGTCGGG 2396
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
320  LysGlyValAlaIleHis..... 324
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2397  TTATTGTTGCGTGTATTGGTTCAGCGCGATTAAAGGATAGCGCGAAGAGTA 2446
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
325  .....LysSerV 327
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2447  TTCGCTCCTATAACTTTATAAAGCCAGTGAAGGAGTCTCGCGCGGGTCT 2496
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
327  aLaspSerGluGlnPheTyrThrAlaMetGluAspAlaTyrPHis..... 341
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2497  TTCAAGCTCTCTTTCACACCGTAGGCGATGTTTTCGACACGACGATCTC 2546
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
342  .....TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2547  CGCTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGCTTTTGGAGACG 2596
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
355  uLeuGluAspSerSerValAsnTyrTrpPheProLysMetArgAspM 372
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:
2597  TTATCTCTATGCCACAGCGGTTTCATCCGTAAAGTACCTGGGTCGTGTGTT 2646
      :|...:|...:|...:|...:|...:|...:|...:|...:|...:

```

372 etValIlePro..... 376  
2647 GTTACATTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696  
377 .....Le 377  
2697 CTTTGATATTTCCAGGAGACACTCCGCTAGGAACCTGAAGATAGTCGCT 2746  
||||| |||||  
377 uPheAspIleSerLeuGluThr.....SerLysArgSerArgL 390  
2747 TGGCGCTCTCTAGGCTATCGTGGAGATTCAATT...AAGTCATG 2793  
GluValLeu.....ValSerLysAspPheValPheThrValLeu 403  
390 ysGluValLeu..... 403  
2794 AAGGCAATGAAGACAGAAGATGGAACCCCTGCCAATTACTGAAGATTC 2843  
404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu.....ThrTyrAlaAs 418  
2844 TGTATATCGCTTCATATGGGGAAGCTTCTTAAGTCCACTGTACGAGGG 2893  
418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430  
2894 CAGGCTCTCTGGCGGTCGAAAGCGACCGTTCGAGTGTTCCTAAG 2943  
431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445  
2944 GGTGTGGTAGCTCGTGGGCTCGCAGCAAGGCCTTTCTGGCATTACGTC 2993  
446 SerLeuLeuGlnSerLeuSerMetThr..... 454  
2994 GTTCTTTTCCACAGGTTCACTATCTTACACCGCGGTTTAACTGAAGATG 3043  
455 .PhePhe..... 456  
3044 AAAGCCTTGATGCTCTGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093  
457 .....LeuHisThrLysLeu 461  
3094 GGCATACTGGAGACGTCGCGGTAGCTGTGAGCAAGTCGTAGTGGAAAC 3143  
462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheSerLeuGlySe 477  
3144 GAAA.....GAATTTTGGAGTGAAGTTTCCTTA.....A 3172  
477 rLysThrValCysGlnHisValTrpAspGluIleSerLeuAlaPheGlyA 494  
3173 ATGACTTCACCACTTTTCGTA.....TTGCGGAATAAGGTGCTTATCGGG 3216  
494 snAlaPheProSerValLysGluArgLeuLeuAsnLysLysLeuIle... 509  
3217 ATATTCTGGGCTCTTTGGGTGCGGCCCCCAATTGCATGGAAGTATAGCG 3266  
509 ..... 509  
3267 CGGAATTCGGGTAAAGCTAAGCTACGCGGCGAGTAGTTACGAACTC 3316  
510 .....ArgAlaAlaGlyAspAlaLeuGluIle. 518  
3317 TAAGTCGTTAAGTTCAACAAGCGCGGTGTTTACGCGGTTTAACTCT 3366  
518 ..... 518  
3367 AGCAGTATCCGTTGGATCTTTAGTCGTGGAAAGAGGTTTTCGTCGGC 3416  
518 ..... 518  
3417 GGTACCGTCACTAGGGCGACCGTAGCTAAACGCTCAAGTCCCTTAGCGT 3466  
519 .....LysValProAspLeuT 524  
3467 TGCTATCGTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGC 3516  
524 yrIleThrPhe..... 527

3517 ATTTGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGG 3566  
527 ..... 527  
3567 GACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCT 3616  
527 ..... 527  
3617 TCTCCAATAATTGGTGCCTGTCTCCGAGGTGTTTGGCGAGGGAAGAGT 3666  
527 ..... 527  
3667 GTCACCTCATTTTACTGCCTATTACGTAGGGGTATCTTTGATCATAAG 3716  
527 ..... 527  
3717 GGGCTTGTTAACGACACCATACCTCAACTTGTACGTCCCAACCGGTAG 3766  
527 ..... 527  
3767 AGGGGAGGAATGTGTACGATGAGACGCTTAGGTATTACGGGACTTTGAC 3816  
528 .....HisAspArgLeuValAlaGluTyrLys..... 536  
3817 TATGACGAAGGTGCTGCTCCATCTCGGACTCAGCATGAAGCGGTTCCCGG 3866  
536 ..... 536  
3867 TCACGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCA 3916  
537 .....SerSerValAspMetProAlaLeuAspIleArgL 548  
3917 CAAATGTGCGGACGTGGGATTAGCACCAACGGGGAAGTTACTGGTGAA 3966  
548 ysArgMet..... 550  
3967 GAAGACACCCATTACCTCGAAGCGTGCATACACTTATGTCGAGGAAGA 4016  
551 GluGluThr.....GluValMetTyrAsnAlaLeuSerGlu... 562  
4017 GGTCCCCCGTCTGCAGCTGTGGCGGAAAGCAAGGTATCCGTCGGGTT 4066  
562 ..... 562  
4067 CTGGTACCGCTGACGCTATGGCTTTTGTGAAAGTGTGAAAAAGGTGC 4116  
563 .....LeuSerValLeuArgGluSerAspLysPheAsp 573  
4117 GACGATGTCTTTCACCAACAGTCTAGTGGGGAACGGCTCGTGAAGTTGA 4166  
574 ValAspValPheSerGlnMet.....CysLysSerLeuG1 585  
4167 GGTGACGCAAGGTTGCTCCCAAGAGCGTCTCGGTGAGGCGCCGA 4216  
585 uValAsp.....ProMetThrAlaAlaLysValIleValAlaValMetS 600  
4217 CACAAGAAAGGGA.....AGAGCTGCAGATGTTAAC 4248  
600 erAsnGluSerGlyLeuThrLeuThrPheGluArgProThrGluAlaAsn 616  
4249 ACAGCACAAACCGGTCACCAAGCGCACAGGGAGCCAGTACAGTCCAG 4298  
617 ValAlaGln...AlaLeuGlnAspGlnGluLysAlaSerGluGlyAlaLe 632  
4299 TCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTACCAGTCCGAGG 4348  
632 uValValThrSerArgGluValGluGluProSerMetLysGlySer.... 647  
4349 TACATGCTCAGAAAGAGTGAACAAGAGTACCATTGGCGACTGTTTCG 4398  
648 .....MetAlaArgGlyGluLeuGlnLeuAlaGlyPheAla 659





```
946 PheProTyrProAlaHisPheSerLysLeuGluValAlaAspGluValGluTh 962
6195 CACTTCTAAATCGTACAGGTGTCCTAGATGTTTGGCTACTGTTGTCT 6244
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
962 rArgArgThrThrLeuArgCysProAlaAspValThrHisTyrLeuAsnA 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6245 CA.....ATGACCGTAAGGGGACGGAAGAAAGTGTAC 6276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
979 rArgTyrGluGlyPheValValSerThrSerValTylsSerVal 995
6277 CQTGAAAGGTCCTTAGCGGTAAAGCACAACACAGTAGTAAGATCGCTGTC 6326
996 SerGlnGluMetValSerGly.....AlaValIleAsnProIleSe 1010
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6327 CAAAGGCGCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACG 6376
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1010 rLysProLeuHisGlyLys..... 1016
6377 TGTACTTTGTGATGACCGAGTGGAGAGTCGGATATGAAGAGTCGTTG 6426
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1017 ..IleLeuThrPheThrGln.....SerAsp...LysGluAlaLeu 1028
6427 AAGCGAAAGGAAAGAACACCACTGATGACATGCATGCAAGCACAGG 6476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1029 LeuSerArgGly...TyrSerGluValHisThrValHisGluValGlnG 1044
6477 AAAACATTCACGTATGTTGTTATTTAGGACGAGAAAGCCGATGACT 6526
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1044 yLuthrThrSerAspValSerLeuValArgLeuThrProThrProIleS 1061
6527 CCCTATTCACT...AAACAACCGCATATCTACTGTTGGTTTTCGAGACAC 6573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1061 crIleAlaGlyAspSerProHisValLeuValAlaLeuSerArgHis 1077
6574 ACACCTCACTGTTTATCCCGCT.....CTCAGCTCAA 6608
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1078 ThrCysSerLeuLysTyrThrThrValValMetAspProLeuValSerI 1094
6609 GTTGACGAT.....AAGTCGGCACATATATTAGCGAC 6642
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1094 cIleArgAspLeuGluLysLeuSerSerTyrLeuLeuAsp 1107
```

seq\_name: SwissProt\_39:RRPO\_TMVB

```
seq_documentation_block:
AC RRPO_TMVB STANDARD; PRT; 1616 AA.
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobacco mosaic virus (strain B935A) (TMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Xue C., Zhou X., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence and genome organization of tobacco
RT mosaic virus isolated from vicia faba.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AJ011933; CAA09875.1; -.
DR EMBL; AJ011933; CAA09874.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR001788; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF00978; RNA_dep_RNAPol2; 1.
DR PFAM; PF01443; Viral_helicase; 1.
DR TRANSFAM; PF01660; Methyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183296 MW; 99A63B441290B6CF CRC64;
```

```
alignment_scores:
  Quality: 285.00      Length: 1879
  Ratio: 0.421        Gaps: 62
Percent Similarity: 36.030 Percent Identity: 16.924

alignment_block:
US-09-301-906-3 x RRPO_TMVB ..
Align seg 1/1 to: RRPO_TMVB from: 1 to: 1616

1207 GCCCTTAATCAAGAGGAGTCTATTTCATCGGTCGTTCCAGTATCACCAA 1256
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    2 AlaTyrThrGlnThrAlaThrThrSerAlaLeuLeuAspThrValArgG 18

1257 TAGACTGCTATTAAAGGACCAATCGCATTCGTTCCCATTTGGACACGA 1306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    18 yAsnAsnSerLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

1307 AATTGTGCGATGATTTTCTCAAAGGAGCGCAATGATTGCGGAAAACCC 1356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    35 laVal...GluGluPheAsnAlaArgAsp.....ArgArgPro 46

1357 TCACATAGTCGCGATGTTCTCTGAAGCGCGGAAAGGAGAGACTGAG 1406
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    47 LysValAsnPheSerLysValIleSerGluGluGlnThrLeuIleAlaTh 63

1407 GGAAGTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTACAGGAGTA 1456
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    63 rArgAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlav 80

1457 GTCACCAATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    80 aHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

1504 TCCAGGAGTGGTAATGTGTGCTTCTTCGATATTGGGGGAGCTTCAC 1553
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

1554 GTATCATGTCAAAGCTGGCCATGTGAACCTGTGTCATGTCATCAATCCAGTCC 1603
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

1604 TAGAGCTTAAAGATGTGAAGCGGAGAGTAATGAG.....ATC 1641
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleG-u 146

1642 CTCCTTCTTTC....ACAGCTGGGAGATTCGTAC..... 1674
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    147 TyrLeuSerLeuArgLeuGluArgGlyGlyLysThrValProAsnPheG 163

1675 ....GTGCCAGTACCTTCTTAACGAAGCGCTTCAAGTCTGTGCTCT 1720
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    163 nLysGluAlaPheAspArgTyrAlaGluIleProGluAspAlaVal.... 178

1721 ACTGTAGTCGAGAAATCGCAGAACTCGGAT..... 1749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```
3520 TGGGCACATGCTCTTCCACGGCAGCTTAATGTTTCTTTGGTTTAGGGAC 3569
490 .....
3570 ATTGCTTGGGGGAGGCTAGCGCAATACATTGGAAGTTTGGAGGCTTCT 3619
491 .....AlaPheG 493
3620 CCAATAATTTGGTGCCTGTTCCTCCAGAGTTTGGCGAGGAAGAGTGTC 3669
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
493 LyAsnAlaPheProSerValLysGluArgLeuLeuAsnArgLysLeuIle 509
3670 ACCTCATTTACTGCCATTACGCTAGGCTAGGCTATCTTGCATCATAGGGG 3719
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
510 Arg.....ValAlaGlyAspAlaLeuGluIleArg.. 519
3720 CTTGCTTAACGACACCATCACTCACTTGCTTACGTCCTCCACCGGTAGAGG 3769
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
520 .....ValProAspLeu...TyrVal..... 525
3770 GGAGGAATGTGTACGATGAGAGCTTAGGTATTACCGGGACTTTGACTAT 3819
525 ..... 525
3820 GACGAAGTGTGCTGCTCATCTGGGACTCAGCATGAAGCGGTTCCCGGTCA 3869
||| ||| : : : : :
526 .....ThrPheHisAspArgLeu..... 531
3870 CGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTACAA 3919
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
532 .....ValThrGluTyrLysAlaSerValA 540
3920 ATGTGCGGCACGTGGGATTAGCACCAACGGGGAAGTTACTGCTGAAGAA 3969
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
540 spMetProAlaLeuAspIleArgLysLysMetGluGluThr..... 553
3970 GAGACCCATTACCTCGAAGCGTCAATACACTTATGTCGAGGAAGAGGT 4019
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
554 .....GluValMetTyrAsnAlaLeuSerGlu..... 562
4020 TGCCCGCTGTCAGCTGTGGCGGAAGACAAAGGTGATCCGCGGTTCTG 4069
562 ..... 562
4070 GTACCGCTGACGCTATGGCTTTTGTGAAGTGTGAAAAAGGTCTCGAC 4119
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
563 .....LeuSerValLeuArgGluSerAspLysPheAspVal 574
4120 GATGCTTTTCCACCAACAGCTCTAGTGGGGAACGGCTCGTIGAGGTTGAGT 4169
||||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
575 AspValPheSerGlnMet.....CysLysSerLeuGluVa 586
4170 GGACGGCAAGGGTTGCTCCCAAGAACGCTGCTCGTGGGCGCCGACAC 4219
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
586 IAsp.....ProMetThrAlaAlaLysValIleValAlaAspMetSerA 601
4220 AAGAAAGGGGA.....AGAGTCGAGATGGTAACACA 4251
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
601 snGluSerGlyLeuThrLeuThrPheGluArgProThrGluAlaAsnVal 617
4252 GCACAAACGGCGTCAACAAAGCGGACAGGAGCCAGCTACAGTCCAGTCT 4301
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
618 Ala...LeuAlaLeuGlnAspGlnGluLysAlaSerGluGlyAlaLeuVa 633
4302 TGTGATTCGCCACAGGCTGATATTCCAAAGGTCAACCGCTCCCGAGGTAC 4351
: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
633 lValThrSerArgGluValGluGluProSerMetLysGlySer..... 647
4352 ATGCTCAGAAAGAGTGAACAAAGTACCATTGGCGACTCTTTCGGGC 4401
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
648 .....MetAlaArgGlyGluLeuLeuGlnLeuAlaGlyLeuAlaGly 660
4402 GCCACGCCAATGCTCGATGAGAAACCCGCCCAAGTGTACGACTCGTGG 4451
```

```
661 AspHisPro..... 663
4452 TGTGAAGATAATTGACAAGGGCAAGCCGCTCGCTCATGTGGCTGAGAAAA 4501
664 .....GluSerSerTyrSerL 669
4502 AACAGGTACAGTCAGCAGCAGCCCAACACAGAGAGTTTGACGATCAATGAA 4551
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
669 ysAsnGluGluIleGluSerLeuGluGlnPheHisMetAlaThrAlaAsp 685
4552 GCCAAGCGCGGTAAACAGCTTTGCATGTTTAGAACGCTGTTCTCGCGGT 4601
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
686 SerLeuIleArgLysGlnMetSerSerIle..... 695
4602 GCAGCTGGATGTTACACGAAGCGGACTATCGCCACCGAGTTCTCAACAG 4651
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
696 .....ValTyrThrGlyProIleLysValGlnGlnMetLysAsn. 708
4652 CATTTACCTTTGTCGATAACTTTGAAAGGAGGAGTCCGCTCTTTTCTCA 4701
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
709 .....PheIleAspSerLeuValAlaSerLeuSerAla..... 719
4702 AAGCTGGGTGAGGGGTATACCTATATATGCTGAGCCATGTTTCATCAGG 4751
719 ..... 719
4752 GTGGCTCGTGCCTAGAGGATATCTTAACGGCAATTAAGTACCCCAAGCG 4801
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
720 .....AlaValSerAsnLeuValLysIleLeuLysAspThrAlaA 733
4802 TCTTCGACCACCTGTTTACTGCAAGAGTACAAGTGGTGGAGCGGTACCA 4851
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
733 laIleAsp.....LeuGluThrArgGlnLysPheGly..... 743
4852 TTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCCTATCTTGAC 4901
743 ..... 743
4902 GGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTCAGAGGGGTG 4951
743 ..... 743
4952 GTAAGTCATGTCATAAACGTAGTTCGGGTCACTATTTTCTTATGCCT 5001
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
744 .....ValLeuaspValAlaSerArgLysTrpLeuIleLysPro 756
5002 TCGGTTTTTCAAAGGACGCACTTGCATTTCAGTAAACTCCATCGACGAAGG 5051
756 ..... 756
5052 GCGCATCAGTTTGACGTTTCAGGGCAACTCGCGCGCTTTTGGTGTAGGCA 5101
756 ..... 756
5102 GGATGTTGCAGTTAGCCGCGCGGTGTCGGATGAGAAGTCACCAAGGTGT 5151
756 ..... 756
5152 CCAAAACCAGAACCCACAGAGCCCAAGGTGTACCAGAACAAATCACACAAA 5201
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
757 .....ThrAla-y 759
5202 ATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAGGGGA 5251
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
759 sSerHisAlaIleTrpGlyValValGluThrHisAlaArgLys..... 772
5252 GGTGCACATACTCGATA.....TGTTGGCAACAAGATTACGTT 5289
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
773 .....TyrHisValAlaLeuLeuGluTyrAspGluGlnGlyValVal 786
5290 AGGAAGTGTGAGTGGCTCAGGCGCTGATAATCCAGTATGGCTCTTGAACC 5339
: : : : : |||
```

6198 TTCTAAATCGGTACAGGTGCGGTTAGATGTTGTGCTACTTGTGTGCTCTCA. 6246  
::: ||||| ||||| ||||| ||||| :::  
963 gargThrThrLeuArgCysProAlaAspValThrHisThrLeuAsnArgA 980  
6247 .....ATGACCGTAAAGGGAACGGAAAGTGTACCCCT 6279  
980 rgTyrgluGlyPheValMetSerThrSerSerValLysLysSerValSer 996  
6280 GAAAAGGTGTTAGCGGTAAAGACAAACCATAGTAGATCGCTGTCCAA 6329  
997 GinGluMetValGlyGly.....AlaAlaValIleAsnProIleSerLy 1011  
6330 AGGCCAATTGGAACCACTGATGAGCTAGCTGAATAAACGCTGACGTGT 6379  
1011 sProSerHisGlyLys.....I 1017  
6380 ACTTGTGCATGACGACAGTGGAGANGTCGGATATGAAAGAGTCTGTTGAG 6429  
1017 leuThrPheThrGln.....SerAsp...LysGluAlaLeuLeu 1029  
6430 GGAAGAGAAAAGAACACACCTGATGACATGTCATGAAGCAGCAGGAAA 6479  
1030 SerArgGly...TyrSerAspValHisThrValHisGluValGinGly 1045  
6480 ACATTCAGTATGATGGTATGTTTGTAGGACGAAGAACCGATGACTCCC 6529  
1045 uThrThrSerAspValSerLeuValArgLeuThrProThrProValSerI 1062  
6530 TATTCACT...AAACAACCGCATATACTTGTGTGTTGCTCGAGACACACA 6576  
1062 leIleAlaGlyaspSerProHisValLeuValAlaLeuSerArgHisThr 1078  
6577 CGCTCACTGTTTATGCGCGCT.....CTGAGCTCAAAAGTT 6611  
1079 CysSerLeuLysTyThrThrValValMetAspProLeuValSerIleI 1095  
6612 GGACGAT...ANGGTGGGCACATATATTAGGCAC 6642  
1095 eArgAspLeuGluLysLysLeuSerSerTyrLeuSerAsp 1107

***This Page Blank (uspto)***

OM of: US-09-301-906-3 to: SPTEMPBL\_15.\* out\_format : pfs  
Date: Jan 31, 2001 4:24 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_n2p.model -DEV=xlp  
-Q/cgcn2\_1/USPTO.spool/US09301906/runat\_31012001\_153037\_9662/app\_query.fasta\_1.6881  
-DB=SPTEMPBL\_15 -OFMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.000 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09301906\_cgcn1\_1354 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-301-906-3  
Query length: 6714  
Database: SPTEMPBL\_15.\*  
Database sequences: 374700  
Database length: 11707915  
Search time (sec): 906.940000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_virus:O71188	+ 3644.00	4224.88	5.4e-228	701	! 071188 grapevine leafroll-associ
sp_virus:Q90604	+ 803.50	923.85	7.0e-44	398	! 09q64 grapevine leafroll-associ
sp_virus:O71209	+ 802.50	902.35	1.8e-43	2473	! 071209 grapevine leafroll-associ
sp_virus:Q90703	+ 746.00	834.08	9.0e-40	3090	! 09q703 beet yellows virus. P34
sp_virus:Q89906	+ 744.50	832.41	1.1e-39	3070	! 089906 beet yellows virus. fus
sp_virus:Q88534	+ 738.50	825.41	2.8e-39	3074	! 008534 sugar beet yellow virus
sp_virus:Q86237	+ 721.00	819.99	5.4e-38	3124	! 06237 citrus tristeza virus.
sp_virus:Q40959	+ 721.00	808.26	3.3e-38	2301	! 040959 little cherry closterov
sp_virus:Q9WID7	+ 712.50	794.97	1.3e-37	3115	! 09wid7 citrus tristeza virus.
sp_virus:Q9IFX0	+ 697.50	777.52	1.3e-36	3115	! 09ifx0 citrus tristeza virus.
sp_virus:O10467	+ 688.00	766.46	5.2e-36	3115	! 010467 citrus tristeza virus.
sp_virus:Q87587	+ 668.00	743.20	1.0e-34	3107	! P87587 citrus tristeza virus.
sp_virus:Q83044	+ 650.00	727.88	1.2e-33	1873	! Q83044 lettuce infectious yell
sp_virus:Q9JGK3	+ 486.00	540.70	4.6e-23	1334	! 09jgk3 soil-borne wheat mosai
sp_virus:Q9WIE3	+ 484.00	538.24	6.2e-23	1350	! 09wie3 chinese wheat mosai
sp_virus:Q9WJD7	+ 484.00	534.69	7.1e-23	1858	! 09wd7 chinese wheat mosai
sp_virus:Q9B08	+ 453.50	502.77	5.9e-21	1345	! 09b08 oat golden stripe virus
sp_virus:Q9IWA9	+ 453.50	499.20	6.7e-21	1853	! 09iwa9 oat golden stripe virus
sp_virus:Q93163	+ 450.50	499.92	9.0e-21	1270	! 093163 beet soil-borne virus.
sp_virus:Q9JGJ7	+ 450.50	499.85	9.0e-21	1277	! 09jg7 sorghum chlorotic spot
sp_virus:Q56790	+ 450.50	496.17	1.0e-20	1778	! 056790 beet soil-borne virus.
sp_virus:Q06359	+ 447.50	495.99	1.4e-20	1320	! 006359 soil-borne wheat mosai
sp_virus:Q89249	+ 447.50	492.37	1.6e-20	1828	! 089249 soil-borne wheat mosai
sp_virus:Q9Q9R0	+ 437.50	484.45	6.3e-20	1308	! 09q9r0 soil-borne rye mosai
sp_virus:Q9JAD4	+ 437.50	480.80	7.3e-20	1816	! 09jad4 soil-borne rye mosai
sp_virus:Q9WK10	+ 436.50	483.30	7.4e-20	1306	! 09wk10 broad bean necrosis vi
sp_virus:Q9YPH5	+ 436.50	479.61	8.5e-20	1820	! 09yph5 broad bean necrosis vi
sp_virus:Q9Q906	+ 427.50	472.81	2.8e-19	1308	! 09q9q6 soil-borne rye mosai
sp_virus:Q9JAD2	+ 427.50	469.15	3.2e-19	1816	! 09jad2 soil-borne rye mosai
sp_virus:Q9QCE8	+ 423.50	468.15	5.1e-19	1308	! 09qce8 european wheat mosai
sp_virus:Q9IWB1	+ 423.50	464.50	5.9e-19	1816	! 09iwb1 european wheat mosai
sp_virus:Q9Q908	+ 422.50	466.99	6.0e-19	1308	! 09q9q8 soil-borne rye mosai
sp_virus:Q9JAD3	+ 422.50	463.33	6.9e-19	1816	! 09jad3 soil-borne rye mosai
sp_virus:Q9QBU0	+ 413.00	455.98	2.5e-18	1302	! 09qbu0 potato mop-top virus. P
sp_virus:Q67704	+ 394.00	430.84	4.7e-17	1707	! 067704 grapevine virus a. puta
sp_virus:Q96806	+ 391.50	427.93	9.8e-17	1707	! 096806 grapevine trichovirus b
sp_virus:Q65853	+ 386.50	431.87	9.9e-17	710	! 065853 beet yellow stunt virus
sp_virus:Q9WKC2	+ 373.00	409.33	9.7e-16	1310	! 09wkc2 beet virus q. hypother
sp_virus:Q90699	+ 373.00	405.69	1.1e-15	1818	! 090699 beet virus q. hypother
sp_virus:Q9IH80	+ 359.00	388.11	9.5e-15	2038	! 09ih80 cherry necrotic rusty m

sp_virus:Q84919	+	354.00	388.37	1.6e-14	1180	! Q84919 pepper ringspot virus
sp_virus:Q84918	+	354.00	384.32	1.8e-14	1698	! Q84918 pepper ringspot virus
sp_virus:P90434	+	338.50	370.25	1.6e-13	1188	! P90434 tobacco rattle virus
sp_virus:P89202	+	337.00	369.06	2.0e-13	1130	! P89202 sunnhemp mosaic virus
sp_virus:O56250	+	336.50	363.89	2.5e-13	1707	! O56250 tobacco rattle virus
seq_name: sp_virus:O71188						
seq_documentation_block:						
ID	O71188	PRELIMINARY;	PRT;	701	AA.	
AC	O71188;					
DT	01-AUG-1998	(TREMBlrel. 07, Created)				
DT	01-AUG-1998	(TREMBlrel. 07, Last sequence update)				
DT	01-JUN-2000	(TREMBlrel. 14, Last annotation update)				
DE	HYPOTHETICAL 77.0 KDA PROTEIN (FRAGMENT).					
OS	grapevine leafroll-associated virus 3.					
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;					
OC	Closterovirus.					
OX	NCBI_TaxID=55951;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=NY1.					
RX	MEDLINE=98264508; PubMed=9603346;					
RA	Ling K.-S., Zhu H.-Y., Drong R.F., Slightom J.L., McPerson J.R.,					
RA	Gonsalves D.;					
RT	"Nucleotide sequence of the 3'-terminal two-thirds of the grapevine					
RT	leafroll-associated virus-3 genome reveals a typical monopartite					
RT	closterovirus."					
RL	J. Gen. Virol. 79:1299-1307(1998).					
DR	EMBL: AF037268; AAC40717.1; "					
DR	INTERPRO: IPR000606; "					
DR	PFAM: PF01443; Viral_helicase1; 1.					
KW	Hypothetical protein.					
FT	NON_TER 1					
SQ	SEQUENCE 701 AA; 77008 MW; 772A764D2CA2A2DC CRC64;					
alignment_scores:						
		Quality: 3644.00	Length: 701			
		Ratio: 5.198	Gaps: 0			
		Percent Similarity: 100.000	Percent Identity: 100.000			
alignment_block:						
US-09-301-906-3 x O71188						
Align seg 1/1 to: O71188 from: 1 to: 701						
4609	GATGCTGATACCAAGGAGGACTATCCGCCACGAGTTCTCAAACGCAATTTAC	4658				
1	AspValtyrAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheTh	17				
4659	CTTTGTCGATAACTTGAAGGGAGGAGTGGCGTCTTTTCTCAAAGCTGG	4708				
17	rPheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuG	34				
4709	GTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGTGGCCT	4758				
34	lyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerGlyTrpPro	50				
4759	CGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCAGCGCTCTTCGA	4808				
51	ArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAs	67				
4809	CCACTGTTTAGTCAGAAGTACAAAGATGGGTGGGAGCGGTACCATTCACG	4858				
67	pHisCysLeuValGlnLysTyrLysMetGlyGlyValProPheHisA	84				
4859	CTGATGACGAGGAGTGTATCCATCAGATAACCCCTATCTTGACGGTCAAT	4908				
84	laAspAspGluGluCysTyrProSerAspAsnProIleLeuThrValAsn	100				
4909	CTCGTGGGGAAGGCAACTTCTCGACTAAGTCAGGAGGAGGTGCTAAGGT	4958				

101 LeuValGlyLysAlaAsnPheSerThrLysCysArgLysGlyLysVa 117  
4959 CATGTCATAAACAGCTTCGGGTGACATATTTTCTATGCTTCGGGT 5008  
117 lMetValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyP 134  
5009 TTCAAAGGACGACTTGATTCAGTAAGTAACACTCCATCGACGAAGGGCGCATC 5058  
134 hcGlnArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIle 150  
5059 AGTTTTCAGCTTCAGGGCAACTCGGCGCTCTTTGCTAGCAGGATGTT 5108  
151 SerLeuThrPheArgAlaThrArgArgValPheGlyValGlyArgMetLe 167  
5109 GCAGTTACCGCGCGCTGTCGGATGACAAGTCCACAGGTGTTCCAAACC 5158  
167 uGlnLeuAlaGlyLysValSerAspGluLysSerProGlyValProAsnG 184  
5159 AGCAACACAGAGCCAGGTGCTACGAGACAATCACACCAAAATTCGGGG 5208  
184 InGlnProGlnSerGlnGlyAlaThrArgThrIleThrProLysSerGly 200  
5209 GCCAAGGCTCTATCTGACGGGAAGTGGTAGGAAGTCAAGGGAGGTCCGAC 5258  
201 GlyLysAlaLeuSerGluGlySerGlyArgGluValLysGlyArgSerTh 217  
5259 ATACTCGATATGCTCGCAACAAGATTACGTTAGGAAGTGTGAGTGGCTCA 5308  
217 rTyrSerIleThrPysGluGlnAspTyrValArgLysCysGluTrpLeuA 234  
5309 GGGCTGATAAATCCAGTGATGCTCTTGAACTGTGACTACACCCCAATGACA 5358  
234 rgAlaAspAsnProValMetAlaLeuGluProAspTyrThrProMetThr 250  
5359 TTTTGAAGTGGTTAAACCGGACCTCTCAAGATGCCGTGCGGAGTACTT 5408  
251 PheGluValValLysThrGlyThrSerGluAspAlaValAlaGluTyrLe 267  
5409 GAAGTATCTGGCTATPAGGCATTGAGAGGACATACAGGGCGTTCGTTATGG 5458  
267 uLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLeuLeuMetA 284  
5459 CTAGAAATATTGCGGTCTACTACCGCGGAGGTGTTCTGAAAGTACCTAAT 5508  
284 laArgAsnIleAlaValThrThrAlaGluGlyValLeuLysValProAsn 300  
5509 CAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGA 5558  
301 GlnValTyrGluSerLeuProGlyPheHisValTyrLysSerGlyThrAs 317  
5559 TCTCATTTTTCATTCAACACAGAGCGGCTTCGCTGTGAGAGACCTACCGT 5608  
317 pleuIlePheHisSerThrGlnAspGlyLeuArgValArgAspLeuProT 334  
5609 ACGTACTCATAGCTGAAAGAGTATCTTTACCAGGGCAAGATGTCGAC 5658  
334 yrValIleAlaGluLysGlyIlePheThrLysGlyLysAspValAsp 350  
5659 CCGGTGTTAGCTTTCGGCGCAACTGTTCTGATCCGACGATATACTGTT 5708  
351 AlaValAlaLeuGlyAspAsnLeuPheValCysAspIleLeuVa 367  
5709 TTTCCACCATGCCATTAAATTTGATAGTGCACCTGAAAGTCCGCTCCATCGC 5758  
367 lPheHisaspAlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysG 384  
5759 GCATGGTGGCGGAATCGTTTAAAGTCCTTCGAATATPAAGTGTATAATGCT 5808  
384 lYMetValGlyGluSerPheLysSerPheGluTyrLysCysTyrAsnAla 400  
5809 CCCCAGGTGGCGTAAACACGACGACGTTAGTGGACGAATTCGTTAAGTC 5858  
401 ProProGlyGlyLysThrThrThrLeuValAspGluPheValLysSe 417

5859 ACCCAATAGCACAGCCACCATTTACGGCTAATGTGGGAAGTTCTGAGGACA 5908  
417 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 434  
5909 TAAATATGGCGGTGAAGAGAGAGATCCGAATTTGGAAGGTCTCAACAGT 5958  
434 leAsnMetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSer 450  
5959 CCTACCACAGTTAACTCCAGGGTGTAAACTTTTATCGTCAGGGGAATGTA 6008  
451 AlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMetTy 467  
6009 TAAAAGGTTTTTGGTGGATGAGTGCACATGATGCATCAAGGCTTACTAC 6058  
467 rLysArgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuG 484  
6059 AACTAGGCGTCTTCGCAACCGCGCTCGGAAGGCTCTTTTTTGGAGAC 6108  
484 InLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAsp 500  
6109 ATAAATCAGATACCATTTCATAAACAGGAGAGGTGTTTAGGATGGATTG 6158  
501 ileAsnGlnIleProPheIleAsnArgGluLysValPheArgMetAspCy 517  
6159 TGCTGTTTTTGTTCCAAAGAAGGCGTTCGTATACACTTCTAAATCGT 6208  
517 salalValPheValProLysLysGluSerValValTyrThrSerLysSerT 534  
6209 ACAGTGTCCGTTAGATGTTTGTCTACTTGTCTCCTCAATGACCGTAAGG 6258  
534 yrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValArg 550  
6259 GGAACGGAAAAAGTGTACCCGTAAGGTCGTTAGCGGTAAGGACAAACC 6308  
551 GlyThrGluLysCysTyrProGluLysValValSerGlyLysAspLysPr 567  
6309 AGTAGTAAGATCGCTGTCCAAAGGCCAATTGGAACCACTGATGACGTAG 6358  
567 oValValArgSerLeuSerLysArgProIleGlyThrThrAspAspValA 584  
6359 CTGAAATAAACGCTGACGTGACTTGTGCATGACCCAGTGGAGAGTCG 6408  
584 laGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSer 600  
6409 CATATCAAGAGTTCGTTGAAGGGAAAAAGAAAGAACACACAGTGTATGAC 6458  
601 AspMetLysArgSerLeuLysGlyLysGlyLysGluThrProValMetTh 617  
6459 AGTGCATGAAGCACAGGGGAAAAACATTTCAGTGCATGCTGGTATTGTTTAGGA 6508  
617 rValHisGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgT 634  
6509 CGAAGAAAGCCGATGACTCCCTATTCTACTAAACACCGCATATACTTGT 6558  
634 hrLysLysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuVal 650  
6559 GGTTTGTCGAGACACACAGCTCAGTGGTTTATGCGCGCTCTGAGCTCAA 6608  
651 GlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLys 667  
6609 GTTGGACGATAAGGTCCGCACATATATTAGCGACGCGTCACCTCAATCAG 6658  
667 sLeuAspAspLysValGlyThrTyrIleSerAspAlaSerProGlnSerV 584  
6659 TATCCGACGCTTGTCTCACAGTTCGCCCGCGCTGGTGTGCTTTCGAGGT 6708  
684 alSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysPheArgGly 700  
6709 ATA 6711  
111  
701 Ile 701



seq\_name: sp\_virus:O906Q4

seq\_documentation\_block:

```

AC O906Q4 PRELIMINARY; PRT; 398 AA.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HELICASE (FRAGMENT).
GN HEL.
OS grapevine leafroll-associated virus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=47985;
RN [1]
RP SEQUENCE FROM N.A.
RA Fazeli C.F., Rezaian M.A.;
RT "Nucleotide sequence and organization of ten open reading frames of
RT the grapevine leafroll-associated virus 1 genome and identification of
RT three subgenomic RNAs."
RL J. Gen. Virol. 0:0-0(2000).
DR EMBL; AF195822; AAF22737.1; -.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicase1.1.
KW Helicase.
FT NON_TER
SQ SEQUENCE 398 AA; 44426 MW; A908109D49C22A23 CRC64;

alignment_scores:
  Quality: 803.50 Length: 394
  Ratio: 2.880 Gaps: 2
Percent Similarity: 70.812 Percent Identity: 44.162

alignment_block:
us-09-301-906-3 x O906Q4 ..
Align seg 1/1 to: O906Q4 from: 1 to: 398

5539 GTTACAAGTCGGGCACAGATCTCATTTTTCATTAACACAGACGGCTT 5588
||||| : : : : : : : : : : : : : : : : : : : : : : : :
3 ValTyrAspAlaAsnPhelysValValAlaAlaAsnAlaProGluAspGlyTh 19

5589 GCCTGTCAGACACCTACCGTACGTACTCATAGCTGAAAGAGTATCTTTA 5638
: : : : : ||||| : : : : : : : : : : : : : : : : : :
19 rThrIlePheAspLeuGluTyrValPheLeuValSerThrGlyThrPheV 36

5639 CCAAGGCGCAAGATGTCGACGGCGTGGTGTGGCGCACAAATCTGTTTC 5688
: : : : : : : : : : : ||||| : : : : : : : : : : : :
36 aIProArgArgAsnLeuGlnAlaValLeuSerArgGlnAspAlaValLeu 52

5689 GTATGCGCAGATATACTGTTTTCACGATGCCATTAAATTTGATAGTGC 5738
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
53 ValCysAspGluLeuLeuValPheHisAspAlaMetAsnLeuArgGlyC 69

5739 ACTGAAGTCGTCGATGGCGATGTCGGCGAATCGTTTAAGCTCTTCG 5788
: : : : : : : : : : : ||||| : : : : : : : : : : : :
69 sValArgLeuAlaLysArgAlaMetValGlyGluTyrMetLysAspVala 86

5789 AATATAAGTCGTATATGTCCTCCCGAGTGGCGGTAAGACGACGCTTA 5838
: : : : : ||||| : : : : : ||||| : : : : : ||||| : :
86 rgIleSerAlaValAsnSerProGlyGlyAsnThrThrArgLeu 102

5839 GTGGACGAATTCGTTAAGTCACCAATAGCACACGCCACCATTCAGGCTAA 5888
||||| : : : : : : : : : : : : : : : : : : : : : :
103 ValAspGluTyrPheGlyArgLysLysArgAlaLysIleAlaAlaAlaAs 119

5889 TGTGGGAAGTCTTGACGACATAAATATGCGGTGAAGACAGAGATCCGA 5938
: : : : : ||||| : : : : : ||||| : : : : : ||||| : :
119 nThrGlySerValAlaAspIleAsnAlaAlaIleArgAlaArgGluGlyL 136

5939 ATTGGAAGTCTCAACAGTGTACACAGTTAACTCCAGGTTGTTAAAC 5988
: : ||| : : : : : ||||| : : : : : ||||| : : : : :

```

```

136 ySLysGluProAspLeuValAlaLysThrAlaAsnSerTTPValIleAsn 152
5989 TTTATCGTCAGGGAAATGTATAAAGGTTTGGTGTGATGAGTGCACAT 6038
||| : : : : : ||||| : : : : : ||||| : : : : :
153 SerHisProArgProAsnSerHisValGlyLeuIleAspGluValTyrMe 169
6039 GATGCACTCAAGGCTTACTCAACTAGCGCTTCGCAACCGCGCGCTCGG 6088
: : : : : ||||| : : : : : ||| : : : : : ||||| : : :
169 tLeuHisLysGlyMetPheGlnLeuThrValValSerMetGlyValLysG 186
6089 AAGGCTCTCTTTTGGAGACATAAATCAGATACATTATCAATAACAGGAG 6138
|| : : : : : ||||| : : : : : ||||| : : : : : |||||
186 luValIleAlaTyrGlyAspLysAsnGlnIleProPheIleAsnArgGlu 202
6139 AAGTGTGTTAGGATGATGCTGTTTGTTCCTCAAGAGAAAGACGCT 6188
||||| : : : : : : : : : : : : : : : : : : : : : :
203 LysThrPheValThrProAsnGluAlaValGluPheAlaGluGluGlnI 219
6189 TGTATACACTCTTAATCGTACAGGTCTCCGTTAGATGTTTGTCTACTTGT 6238
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : :
219 eAspTyrThrAspIleSerTyrArgCysProAlaAspValCysTyrVal 236
6239 TGTCTCAATGACCTAAGGGGAACGGAAGTGTACCTGAAAGAGGTC 6288
||||| : : : : : : : : : : : : : : : : : : : : : :
236 euSerSerMetThrAspAlaArgGlyLysLysMetTyrProAsnGlyVal 252
6289 GTTAGCGGTAAGGACAAACCAAGTAGTAGCTGTCTCAAAAGGCCAAT 6338
: : : : : ||| : : : : : ||||| : : : : : ||||| : : :
253 PheProAsnGlyAspValArgProLeuArgSerPheGluLysValProI 269
6339 TGAACCACTGATCAGCTAGCTAGCTGAAATAAACGCTGACGTCTATTGCA 6388
||||| : : : : : : : : : : : : : : : : : : : : : :
269 eAlaThrProGluAspAlaLeuLeuTyrGluAlaAspValTyrLeuThrM 286
6389 TGACCCAGTTGGAGAGTCGGATATGAAGAGTCTGTTGAAGGGA. .... 6432
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
286 eThrGlnAsnGluLysAlaGluMetGlnLysAlaValAlaLysMetGlu 302
6433 .....AAAGAAAGAAACACCA...GTGATGACAGTGCATGAAGCACA 6473
||||| : : : : : ||| : : : : : ||||| : : : : : |||||
303 ValValAlaGlyLysLysArgProAspValIleThrHisGluAlaGlu 319
6474 GGAAACAACTTCAGTCATGCTGTTTGTAGGACGAAAGAAACCGCATG 6523
||||| : : : : : ||||| : : : : : ||| : : : : : |||||
319 nGlyLysThrTyrGluAsnValValLeuValArgLeuLysLysAlaAspA 336
6524 ACTCCCTATTCTACTAAACACCGCATATCTGTTGTTGTTGTCGAGACAC 6573
|| : : : : : ||||| : : : : : ||||| : : : : : |||||
336 sProIlePheSerArgLysProHisIleValAlaAlaLeuSerArgHis 352
6574 ACACGCTCACTGTTTATGCCGCTCTGAGCTCAAGTTGGACGATAAGGT 6623
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
353 ThrArgSerMetLysTyrAlaValLeuSerSerLysMetThrAspThrI 369
6624 CGGCACATATATTAGGACCGCTCACCCTCAATCAGTATCCGACGCTTTC 6673
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : :
369 eSerLysLeuIleAspGlyThrSerAlaGlyLysValSerAspValLeuL 386
6674 TTCACAGTTCGCCCGCGTGTGTTGCTTCCA 6705
|| : : : : : |||||
386 euGlnGlnLeuGlnArgAsnAspArgPheArg 396

seq_name: sp_virus:O71209
seq_documentation_block:
ID O71209 PRELIMINARY; PRT; 2473 AA.
AC O71209;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE METHYLTRANSFERASE/HELICASE POLYPROTEIN (FRAGMENT).
OS grapevine leafroll-associated virus 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;

```

```

OC Closterovirus.
OX NCBI_TaxID-84003;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98264507; PubMed-9603345;
RA Zhu H.Y., Ling K.S., Gosczyński D.E., McPerson J.R., Gonsalves D.;
RT "Nucleotide sequence and genome organization of grapevine leafroll-
associated virus-2 are similar to beet yellows virus, the
RT closterovirus type member."
RL J. Gen. Virol. 79:1289-1298(1998).
DR EMBL; AF039204; AAC40855.1; .
DR INTERPRO; IPR000606; .
DR INTERPRO; IPR002588; .
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Methyltransferase; 1.
KW Transferase; Methyltransferase; Helicase; Polyprotein.
FT NON_TER 1
SQ SEQUENCE 2473 AA; 277096 MW; F8DEBBA6A58ED425 CRC64;

alignment_scores:
  Quality: 802.50      Length: 2475
  Ratio: 0.681        Gaps: 115
Percent Similarity: 47.596 Percent Identity: 22.424

alignment_block:
US-09-301-906-3 x 071209 ..

Align seg 1/1 to: 071209 from: 1 to: 2473

454 CAAGATGACGACAGCCCAACGGCAAGGTCACATTTGGCTGTAAAGCC 503
|||||: ||| |||||: |||||: |||||: |||||:
158 GlnSerSerValLysThrLysArgLysArgLysAlaHisArgThrValGluG1 174
504 AACAGCTGTGGCGCCAGCTACCACTTCCATAAAACAGGAAGCACTGG 553
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
174 uArgAlaAlaGlyGlySerValGluGlnProArgGlnLysArgLysLeuAasp 191
554 AGCCA...GCCAATCAGTCCACACACAGTCGTG. .... 585
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
191 LuLysGlyCysGlyArgValProSerGlyGlyPheSerHisLeuLeuVal 207
586 .....GAGCAAGGCCCTTGACGTTGGCTTTTCTTCAAG 623
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
208 GlyAsnLeuAsnGluValArgArgLysValAlaAlaGlyLeuLeuArgPh 224
624 TAAAGTGGGGTGATGAGAGCGCTGTCATCTTCGCGAAAGGAAAT 673
|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
224 eArgValGlyGlyAspMetAsp. .... 231
674 TGTTTACAGGCCCTTAATGTTCTTATTTGATGTAAGAACAACGTTGCTT 723
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
232 ..PheHisArgSerPheSerThrGlnAlaGly...HisArgLeuLeuVal 246
724 TGGGCTAA.....ATCTGGGATGAACCTC 749
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
247 TrpArgArgSerSerArgSerValCysLeuGluLeuTyrSerProSerLy 263
750 TCGTAGGAGAGGTTATTTTAGTCAAA.....GATAGAGCTGTA 790
|||||: |||: |||: |||||: |||||: |||||: |||||: |||||:
263 sAsnPheLeuArgTyrAspValLeuProCysSerGlyAspTyrAlaAlam 280
791 AATTCTCCCTATTGTGGGGTAGGGCTACGATCGAGGACTTTCATCGTG 840
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
280 etPheSerPheAlaAlaGlyArgPheProLeu..... 291
841 AATACAGCCCAAGGTGTGATGTTGCTTGGCCGCGCATGACTTGTGGAG 890
|||||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
292 .....ValLeuMetThrArgIleArgTyr..... 299
891 TATGCGGAAGGGCGTTGTATGACCAACCAAGGGTGGTGTGTTA 940
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
300 .....ProAsnGlyPheCysTyrLeuA 307

```

```

941 ACAATGAGAGGCTGAGGGGAGAAATTTACAGACGTCGTTGCTTCTCATCT 990
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
307 laHisCysArgTyrAlaCysAlaPheLeuLeuArgGlyPheAspProLys 323
991 TCCTTTTCATAGGT.....TT 1007
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
324 ArgPheAspIleGlyAlaPheProThrAlaAlaLysLeuArgAsnArgMe 340
1008 CTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTTGGGGCA 1057
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
340 tValSerGluLeuGlyGluArgSerLeu..... 349
1058 CGAACATACTACATGCCATCACTCAATGAAGAGCGTACCTTTGGTGG 1107
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
350 .....GlyLeuAsnLeuTyrGlyAlaTyrThrSer 359
1108 AAGGCG.....GGAGACGCTCTA 1124
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
360 ArgGlyValPheHisCysAspTyrAspAlaLysPheIleLysAspLeuA 376
1125 TCTCCCAATGTCCCAAAACCGCTATCGCTGGCGATAGGACACGCT 1174
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
376 glu.....MetSerAlaValIleAlaGly...LysAspGlyV 388
1175 TGGGAGGGAGATCTTGGCTCCGTCGCAATGCCCTTAATCAAGAGAG 1224
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
388 alGluGluValValProSerAspIleThrProAlaMetLysGlnLys... 403
1225 GTCTATTATCGTCGTTTCGAGTATCACCATACTAGCTGGTATTAGGGA 1274
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
404 .....ThrIleGluAlaValTyrAspArgLeuTyrGlyGlyTh 416
1275 CCAATCGGCATTGCTTCCCATTTGGACACGAAATTTGCGGATATGTTT 1324
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
416 rAspSerLeuLeuLysLeuSerIleGluLysAspLeuIleAspPheLysA 433
1325 CTCAAAGGACCAATGATTCGCGAAAAACCTCACATAGTCCGCGATGTG 1374
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
433 snAspValGlnSerLeuLysLysAspArgProIleValLysValProPhe 449
1375 TTTCTCAACCCCGGGAAGGAGAGAGCTGAGGAACTCTTCCACAGCT 1424
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
450 TyrMetSerGluAlaThrGlnAsnSerLeuThrArgPheTyrProGlnPh 466
1425 TTCGATACAGTTCTCCGACTCGGTGAGGAGTAGTCACCCATTCGCTAATG 1474
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
466 eGluLeuLysPheSerHisSerHisSerHisProAlaAlaAlaA 483
1475 CCATCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGGTAAATGTG 1524
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
483 laSerArgLeuLeuGluAsnGluThrLeuValArgLeuCysGlyAsn... 498
1525 TCCTTCTCGATATTGGGGGAGCTTTCAGTATCATGTCAAAGCT...CG 1571
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
499 SerValSerAspIleGlyCysProLeuPheHisLeuHisSerLysTh 515
1572 CCATGTGAACCTGCTATGCAATCCAGTCTAGACGTTAAAGATGTGA 1621
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
515 rGlnArgArgValHisValCysArgProValLeuAspGlyLysAspAlaG 532
1622 AGCGGAGA...ATCATGAGATCCTCTTTTCCACAGCTGGGGAGAT 1668
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
532 lnArgArgValValArgAspLeuGlnTyrSerAsnValArgLeuGlyAsp 548
1669 TCGTAGTCTCCAGTACCTTCTTAAGCGGCTTCAAAGTCTGTCTC 1718
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
549 .....AspAspLysIleLeuGlu...GlyProArgAsnIleAs 560
1719 TTAGTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGTT 1768
|||: |||: |||||: |||||: |||||: |||||: |||||: |||||:
560 pIleCysHisTyrProLeuGlyAlaCysAspHisGluSerSerAlaMetM 577

```

```

1769 TTATGGTGGATGTACGATATATCCCGCAGCGAGGTAGCAGAGGCTATG 1818
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 eMetValGlnValTyraSPalaSerLeuTyrgluileCysGlyAlaMet 593
1819 GATAAGAAGGTCGCGTGGTTTCGACATAGCTCTTATGTTCCCGCTGGA 1868
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
594 ileLysLysSerArgIleThrTyrluThrMetValThrProGlyGI 610
1869 GTTGTGTAGCGTAACGGTGAAGTTTACTTGGGAAGACGTGATACGTTGG 1918
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 uPheLeuAspGlyArgGluCysValTyrlMetGluSerLeuAspCysGluI 627
1919 TGAAGAGGGA.....GGTGATTACTGCCTTACAATGTGTGTCAGTGT 1962
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
627 leGluValAspValHISAlaAspValValMetTyrlLysPheGly..... 641
1963 GGTGAGATGATGAACATTCCTCTCTAAGCTAAGCGGGTTCACCTT 2012
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
642 SerSerCysTyrlSerHisLysLeuSerIleleLysAspIleMetThrTh 658
2013 TTCTTATGACGCATCGTCCGGGAACGTGTTTAACTAGATATGAGG 2062
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
658 rProTyrlLeu...ThrLeuGlyGlyPheLeuPheSerValGluMetTyrg 674
2063 GATACCGTTGCTGTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCA 2112
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
674 luValArgMetGlyValAsnTyrlPheLysIleThrLysSerGluValSer 690
2113 CTGGAACTGAGGTACG.....TATAGGTCGGTG..... 2142
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
691 ProSerIleSerCysThrLysLeuLeuArgTyrlArgArgAlaAsnSerAs 707
2143 .....GTCCGTCGTTCTGGGCAAACTCGTGGTTCATCA 2176
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
707 pValValLysValLysLeuProArgPheAspLysLysArgArgMetCysL 724
2177 TACCTGTTGTAGCTGTTCTAGTGTCTCTTTAAGCAATAGTCTCGAT 2226
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
724 euPro.....GlyTyrlAspThrIleTyrlLeuAsp 733
2227 TCGGACTTTGTCACAGGATCTATTCTACGCGCTCAACACTATAGGAC 2276
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
734 SerLysPheValSerArgValPheAspTyrlValValCysAsnSerAl 750
2277 ATTGAGAATAGAACGTTTGATGATCGGTTGGGCGGTGAGTCGCCAA 2326
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
750 aValAsnSerLysThrPheGluTrpValTrpSerPheIleLysSerL 767
2327 AGACCATGTCATTACAGGAGTCGCGTGTCTCCACAGCAAGTTGATTT 2376
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
767 ysSerArgValIleIleSerGlyLysIleIleHisLysAspValAsnLeu 783
2377 TCTCCTGATGATGTGGGTTTAGTGTGCGTGTATGCTCAGCGCAT 2426
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
784 AspLeuLysTyrlValGluSerPheAlaAlaValMetLeuAlaSerGlyVa 800
2427 TAAGGATAGG.....GGGAAGATATTTCGTCCTCAT 2458
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
800 lArgSerArgLeuAlaSerGluTyrlLeuAlaLysAsnLeuSerHisPheS 817
2459 ACTTTATAAAGCCAGTAGGGGAGTCTCCCGGGTCTTCAAGCTCTTC 2508
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
817 er..... 817
2509 TTTCAGACCGTAGCGGATGTT...TTTTCGACGCGAGTCTCGCTATGC 2555
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
818 .....GlyAspCysSerPheIleGluAlaThrSerPheValle 830
2556 TAAGCAATGGTGCACGAT.....AACTTCAACGTTTGGAGACGC 2596
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
830 uArgGluLysIleArgAsnMetThrLeuAsnPheAsn.....GluArgL 845
2597 TTATGTCTATGCCAGA.....GCCTTCATCGTAAAGTACCTGGGTCT 2640

```

```

845 euLeuGlnLeuValLysArgValAlaPheAlaThrLeuAspValSerPhe 861
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2641 GTTGTGTTTACCATTTTCACATCTCTGGAGCTTCAGACAGGTTGGAGCTCA 2690
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
862 LeuAspLeuAspSerThrLeuGluSerIleThrAspPheAlaGluCysLy 878
2691 GGGTGCCTTTGATATTTCGAAG.....GAGACCTTCG 2722
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
878 sValAlaIleGluLeuAspGluLeuGlyCysLeuArgAlaGluAlaGluA 895
2723 GTAGGAACCTCAACAATAGTCGCTTGGCGCTCTCTCTAGGGCTATCGTG 2772
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
895 snGluLysIleArgAsn.....LeuAla 902
2773 GAAGATTCAATTAAAGTCAATGAAGCAATGAAGACAGAA.....GA 2813
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
903 GlyAspSerIleAlaAlaLysLeuAlaSerGluIleValValAspIleAs 919
2814 TGGAAACCCCTGCCAATTTACTGAAGATTCGTATATGCGTTTCATAATGG 2863
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
919 pSerLysProSerProLysGln.....ValG 928
2864 GGAACGTTTCTAACCTCCACTGT.....ACGAGGCGAGGT 2898
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
928 lyAsnSerSerSerGluAsnAlaAspLysArgGluValGlnArgProGly 944
2899 CTCTTTGGCGTTTCGAAGACCGCTGTTTCGAGTGTTCCTAAAGGTTT 2948
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
945 LeuArgGlyGlySerArgAsnGlyValValGlyGlu..... 956
2949 GGTAGCTGTCGGGCTCCGACGAAGCGCTTTTCGGCATTTACGTCGTCT 2998
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
957 .....PheLeuHisPheV 961
2999 TTTCACAGGTTTCACTATTCTACGACCGCGTTTAACTGAAGATGAAGG 3048
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
961 alValAspSerAlaLeuArgLeuPheLysTyrlAlaThrAspGlnGlnArg 977
3049 CTGTGATGCTCTGTCGCGACAGAGATGCTATAAACTCACCGGTGGCAT 3098
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
978 ileLysSerTyrlValArg.....PheLeuAspSerAlaValSerPh 991
3099 ACTGGAGACGTCGCGCTAGCTGTGACAAAGTCGTAGCTGGAACGAAG 3148
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
991 eLeuAspTyrlAsnTyrlAspAsnLeuSerPheIleLeuArgValLeuSerG 1008
3149 AATTTTCGAGT..... 3159
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1008 luGlyTyrlSerCysMetPheAlaPheLeuAlaAsnArgGlyAspLeuSer 1024
3160 .....GAAGTTTCTTAAATGA 3176
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1025 SerArgValArgSerAlaValCysAlaValLysGluValAlaThrSerCy 1041
3177 CTTCACACACTTTTCGTATTCGGAATAAGTCTTATCGGATATTCGTGG 3226
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1041 sAlaAsnAlaSerValSerLysAlaLysValMetIle..... 1053
3227 CGTCTTTGGGTGGCGCCCAATTCATGGAAGTATAGCGCG.....GGAATT 3273
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1054 ..ThrPheAlaAlaValAlCysAlaMetMetPheAsnSerCysGlyPhe 1069
3274 GCGGCTAACGCTAGAAGGTACGCGGC..... 3300
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1070 SerGlyAspGlyArgGluTyrlLysSerTyrlIleHisArgTyrlThrGlnVa 1086
3301 .....AGTAGTTACGAACTCTAAGCT 3322
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1086 lLeuPheAspThrIlePhePheGluAspSerSerTyrlLeuProIleGluV 1103
3323 CGTTAGTTTCACAAGCGCGCGGTGTTTACGCGGTTTAACTCTTAGC... 3369
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

4836 .....GGGTGGAGCGGTACCATTC 4855
1651 ValArgLeuAlaSerAlaLeuLeuSerLeuGlyTrpGluGlyLeu..... 1665
4856 ACGTGATGACGAGAGTGTATCCATCAGATAACCCCTATC...TTGACG 4902
1666 .....LysLeuLeuLeuSer..HisHisAsnLeuLeuPheLeuCys 1678
4903 GTCATCTCGTGGGAAGGCAAACTTCCTCAGCTAAAGTCGAGGAAGGCTGG 4952
1679 PheAlaLeuValAspAspValAsnValLeuIleLysValLeuGlyGlyLe 1695
4953 TAAGTCATGGTCATAAACCTAGCTCG..... 4980
1695 userPhePheValGlnProIlePheSerLeuPheAlaAlaMetLeuLeuG 1712
4981 .....GGTGACTAT..... 4989
1712 InProAspArgPheValGluTySerGluLysLeuValThrAlaPheGlu 1728
4990 TTCTTTATGCTTGC.....GGTTTTC 5012
1729 PhePheLeuLysCysSerProArgAlaProAlaLeuLeuLysGlyPhe.. 1744
5013 AAGGACGCACCTTCATTCATCACTAACTCCATCGACGAGGCGCATCAGTT 5062
1745 .....PheGluCysValAlaAlaAsnSerThrValSerL 1755
5063 TGACGTTTCAGGCAACTCGCGCGCTCTT.....GGT 5094
1755 ysthrValArgArgLeuLeuArgCysPheValLysMetLeuLysLeuArg 1771
5095 GTAGGCAGGATGTTGACGTAGCCGCG....GGCGTGTGGATGAGAAGTC 5141
1772 LysGlyArgGlyLeuArgAlaAspGlyArgGlyLeuHisArgGlnLysAl 1788
5142 A...CCAGGTGTTCCAAACAGCACACAGCCAGCCAGGTGCTACCGAA 5188
1788 aValProValIleProSerAsnArgValThrAspGlyValGluArgL 1805
5189 CAATCACACCAAAATCGGGGGCAAGGCTCTATCT..... 5223
1805 euserValLysMetGlnGlyValGluAlaLeuArgThrGluLeuArgIle 1821
5224 .....GAGGGAAGTGGTAGGAAGTCAA 5246
1822 LeuGluAspLeuAspSerAlaValIleGluLysLeuAsnArgArg..Ar 1837
5247 GGGGAGTGCACATCTCGATATGGTCGCAACAGATTACCTTAGGAG. 5295
1837 gAsnArgAspThrAsn.....AspAspGluPheThrArgProA 1850
5296 .....TGTGAGTGGCTCAGGGCT 5313
1850 laHisGluGlnMetGlnGluValThrThrPheCysSer.....LysAla 1864
5314 GATAATCCAGTATGGCTCTTGAA..... 5337
1865 AsnSerAlaGlyLeuAlaLeuGluArgAlaValLeuValGluAspAlaI 1881
5338 .....CCTGACTACACCCCATGACATTTGAAGTGGTTAAACCGGG. 5379
1881 eLysSerGluLysLeuSerLysThrValAsnGluMetValArgLysGlys 1898
5380 .....ACCTCTGAAGATGCGCTCGTG.....GAGTACTTG 5409
1898 erThrThrSerGluValAlaValAlaLeuSerAspAspGluAlaVal 1914
5410 AAGTATCTGGCTATAGGCATTGAGAG.....ACATACAG 5444
1915 GluGluIleSerValAlaAspGluArgAspAspSerProLysThrValAr 1931
5445 G.....GCGTTGCTTATGGCTAGAA 5464

```

```

1931 gIleSerGluTyLeuAsnArgLeuAsnSerSerPheGluPheProLysP 1948
5465 ATATTGCGCTACCTACCGCGAAGGTGTTCTGAAGTACCTAATCAAGTT 5514
1948 rolleValValAspAsnLysAspThrGlyGlyLeuThrAsnAlaVal 1964
5515 TATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCAT 5564
1965 ArgGlu.....PheTyTyMetGlnGluLeuAlaLeuPheG 1977
5565 TTTTCATTCAACA.....CAAGACGGCTTCGCTGTG..... 5595
1977 uIleHisSerLysLeuCysThrTyTyArgPglNLeuArgIleValAsnP 1994
5596 .....AGAGACCTACCGCTACGTACTCAT 5619
1994 heAspArgSerValAlaProCysSerGluAspAlaGlnLeuTyValArg 2010
5620 GCTGAAAGGTATCTTTTACCAGGGCAAGATGTCGACGCG...GTGGT 5666
2011 LysAsnGlySerThrIleValGlnGlyLysGluValArgLeuHisIleLy 2027
5667 AGCTTTGGCGCACAACTCTGTCGTATGCGACGATATACTG..... 5706
2027 sAspPheHisAspHisAspPheLeuPheAspGlyLysIleSerIleAsnL 2044
5707 .....GTTTTCACGATGCCATTAATTGTGATA 5733
2044 ysArgArgArgGlyGlyAsnValLeuTyHisasp.....AsnLeu... 2057
5734 GTGTCATGAAAGTCGTCGATCGGCATGGTGGCGAATCGTTT..... 5778
2058 ...AlaPheLeuAlaSerAsnLeuPheLeuAlaGlyTyTyPropheSerAr 2073
5779 .....AAGTCCTTCGAATATAAGTGCCTATTAATGCTC 5809
2073 gSerPheValPheThrAsnSerSerValAspIleLeuLeuTyGluAlap 2090
5810 CCCAGGTGGCGTTAAGACACGACGCTTAGTGGACGAATTCGTT..... 5853
2090 roProGlyGlyGlyLysThrThrLeuIleAspSerPheLeuLysVal 2106
5854 .....AAGTCACCCAATAGCACACCCACCATACGGCTAATGTGGGAAG 5897
2107 PheLysLysGlyGlyValSerThrMetIleLeuThrAlaAsnLysSerSe 2123
5898 TTCTGACGACATAAATATGCGGTGAAGAGAGAGATCCGNAATTGGAA. 5946
2123 rGlnValGluIleLeuLysLysValGluLysGluValSerAsnIleGluC 2140
5947 .....GGTCTCAACAGTGTACCACAGTT 5970
2140 ysGlnLysArgLysAspLysArgSerProLysLysSerIleTyThrIle 2156
5971 AACTCCAGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAGGGTTT 6020
2157 AspAlaTyTyLeuMetHisHis.....ArgGlyCysAspAlaAspValLe 2171
6021 G...GTGGATGAGTGCACATGATGCATCAAGGCTTACTACAATAGCGG 6067
2171 uPheIleAspGluCysPheMetValHisAlaGlySerValLeuAlaCysI 2188
6068 TCTTCGCAACCGCGCTCGAAGGCGCTCTTTTGGAGACATAAATCAG 6117
2188 leGluPheThrArgCysHisLysValMetIlePheGlyAspSerArgGln 2204
6118 ATACCATTCATAACACAGG.....GAGAAGGTGTTTAGGATGGATTG 6158
2205 IleHisTyIleGluArgAsnGluLeuAspLysCysLysLeuTyGlyAspLe 2221
6159 TGCTGTTTTTGTCCAAAGGAAGGAAAGCGGTGTATACACTTCTAAATCGT 6208

```

2221 uAspArgPheVal...AspLeuGlnCysArgValTyrGlyAsnIleSerT 2237  
6209 ACAGGTGTCGGTAGACTGTTCTACTACTGTTCTCTCTCAATG..... 6249  
||||| 6210  
2237 YrArgCysProTrpAspValCysAlaTrpLeuSerThrValTyrGlyAsn 2253  
6250 .....ACGTAAGGGGACGGAAGAAAGTGTACCTGAAAGGTGCT 6290  
||||| 6251  
2254 LeuIleAlaThrValTyrGlyGluSerGlu..... 2263  
6291 TAGCGGTAGGACAAACAGTAGTAGTCCGCTGCTCCAAAGGCCCAATTG 6340  
2264 ....GlyLysSer.....SerMetArgIleAsnGluIleA 2274  
6341 GAACCACTGATGACGTAGTGAATAAAGCGTACGCTGTACTTGTCCATG 6390  
..... 6391  
2274 snSerValAspAspLeuValProAspValGlySerThrPheLeuCysMet 2290  
6391 ACCCAGTGGAGAGTCGATATGAGAGGTGCTTGAAGGGAAGAAGGA.. 6438  
||| 6439  
2291 LeuGlnSerGluLysLeuGluIleSerLysHisPheIleArgLysGlyLe 2307  
6439 ..AAGAAACACACAGTCATGACATGATGAGAGGTGCTTGAAGGGAAGAAGGA.. 6438  
||| 6439  
2307 uThrLysLeuAsnValLeuThrValHisGluAlaGlnGlyGluThrTyrA 2324  
6488 GTATCTGTGATGTTAGTGGACGACGAGCGGACGCTCCCTATTTCAC 6537  
:: ||| 6538  
2324 IaArgValAsnLeuValArgLeuLysPheGlnGluAspLupProPheLys 2340  
6538 AAACACCGCATATCTGTTGGTTGCGGACACACACGCTCCTGCTGCT 6587  
:: ||| 6588  
2341 SerIleArgHisIleThrValAlaLeuSerArgHisThrAspSerLeuTh 2357  
6588 TTATCCGCTCTGAGCTCAAACTGCGACATAAGCTCGGCACATATATTA 6637  
2357 rTyrAsnValLeuAlaAlaArgArgGlyAspAlaThrCysAspAlaIleG 2374  
6638 GCGACGCGCTCA 6648  
:: ||| 6649  
2374 InLysAlaAla 2377

seq\_name: sp\_virus:Q9Q703

seq\_documentation\_block:  
ID Q9Q703 PRELIMINARY; PRT; 3090 AA.  
AC Q9Q703;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE P348.  
OS Beet yellows virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;  
OC Closterovirus.  
OX NCBI\_TaxID=12161;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BVY-4;  
RX MEDLINE=20079557; PubMed=10611288;  
RA Peremyshlov V.V., Hagiwara Y., Dolja V.V.;  
RT "HSP70 homolog functions in cell-to-cell movement of a plant virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BVY-4;  
RA Peremyshlov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;  
RL Submitted (Sep-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF190581; AAF14300.1; -;  
DR INTERPRO; IPR000606; -;  
DR INTERPRO; IPR001788; -;  
DR INTERPRO; IPR002588; -;  
DR PFAM; PF00978; RNA\_dep\_RNapol12; 1.  
DR PFAM; PF01443; Viral\_helicase1; 1.

DR PFAM; PF01660; Vmethyltransf; 1.  
SQ SEQUENCE 3090 AA; 347990 MW; 09C779A57DE874B9 CRC64;  
alignment\_scores:  
Quality: 748.00 Length: 2197  
Ratio: 0.740 Gaps: 97  
Percent Similarity: 46.017 Percent Identity: 22.531  
alignment\_block:  
US-09-301-906-3 x Q9Q703 ..  
Align seg 1/1 to: Q9Q703 from: 1 to: 3090  
1222 GAGGTCTATTTCATCGGTC.....GTTTCGAGTATCACCATAAG 1259  
||||| 1260  
596 GluIleThrSerSerLeuLysHisLysAlaIleGluSerValTyrGluAr 612  
1260 ACTGTTATTAAAGGACCAATCGGCATTTGCTTTCCCATTTGGACACGAAAT 1309  
||||| 1310  
612 gValSerIleHisLysAspAsnLeuLeuAlaArgSerValGluLysAspL 629  
1310 TGTGCGATATGTTTCTCAAAGGCGCAATGATTCGCCGAAAAACCTCA 1359  
|| ||||| 629  
629 euIleAspPheLysAspGluIleLysSerLeuSerLysGluLysArgSer 645  
1360 CATAGTCGCGATGTGTTCTGAAGCGCGGAAAGGAGGAGCTGAGGGA 1409  
||||| 1410  
646 ValThrValProPheTyrMetGlyGluAlaValGlnSerGlyLeuThrAr 662  
1410 ACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGCTCAGGAGTAGTC 1459  
: ||||| 662  
662 gAlaTyrProGlnPheAsnLeuSerPheThrHisSerValTyrSerAsph 679  
1460 ACCCATTCCTTAATGCCATCGCGAGCTGTTCAATGGAATCTTTTCCAGG 1509  
||| ||| 679  
679 isProAlaAlaAlaGlySerArgLeuLeuGluAsnGluThrLeuAlaSer 695  
1510 AGGTGTGTAATGTGTGCTTCTTCGATATTGGGGGAGCTTCACGTTATCA 1559  
::: 696  
696 MetAlaLysSer...SerPheSerAspIleGlyGlyCysProLeuPheHl 711  
1560 TGTCAAAGCTGCCATGTGAATGTGATGCAATCCAGTCCCTAGACG 1609  
||:|||| ||| 711  
711 sIleLysArgGlySerThrAspTyrHisValCysArgProIleTyrAspM 728  
1610 TTAAGATGTGAAGCGGAGAAATCAATGAGATCTCTTTTCCACAGCT 1659  
:||||| 728  
728 etLysAspAlaGlnArgValSerArg.....GluLeuGlnAlaArg 742  
1660 GGGGAGATTCGTACGTGCCAGTGACCTTCACTGAAGCGGCTTCAAA 1709  
||| 743  
743 GlyLeuValGluAsnLeuSerArgGluGlnLeuValGluAlaGlnAlaAr 759  
1710 GTCTGTCTTACTGTAGTCGAGAATCGCAGAACTCGATCTTAGACCG 1759  
: ||||| ||| 759  
759 g...ValSerValCysProHisThrLeuGlyAsnCysAsnValLysSerA 775  
1760 ATCGCGGTTTTATGTTGGATGTGTACGATATATCCCGCAGCAGGTAGCA 1809  
||:| 775  
775 spValLeuIleMetValGlnValTyrAspAlaSerLeuAsnGluIleAla 791  
1810 GAGGCTATGATAAAGAGGCTCGCTGTTTCGACATAGCTTATGTT 1859  
:::||||| 792  
792 SerAlaMetValLeuLysGluSerLysValAlaTyrLeuThrMetValTh 808  
1860 CCCCTGGAGTTGTTGTACGGTGAACGGTGAAGTTTACTTTGGAAGAATC. 1908  
||| 808  
808 rProGlyGluLeuLeuAspGluArgGluAlaPheAlaIleAspAlaLeuG 825  
1909 .....GATACGTTGGTGAAGAGGAGGAGGTGATTACCTGCGCTACAATGTT 1953  
||:||||| 825  
825 lyCysAspValValAspThrArgArgAspMetValGlnTyrLysPhe 841



```

1364 GlyGluHisGlyIlePheThrLysValPheLeuLeuCysGlyPheLeuPr 1380
      |||:|||||
3516 CATTGG .....GCACGCTCTCCACGGCCTTAATGCT 3550
      |||:|||||
1380 oValPheValArgLysCysValAlaLeuCysValProGlyAspMetAlaT 1397
      |||:|||||
3551 TTTTC .....TTGGTTTAGGACATTCCTGGCGGAGG 3585
      |||:|||||
1397 hrTyAlaArgPheLeuGluTyGlyValAspLeuPhePheLeuGly 1413
      |||:|||||
3586 GCTACGGCAATACATTGAAGTTTGAGGCTTCTCCAATAATTGGTGCGC 3635
      |||:|||||
1414 ArgSerValAsnSerIleLys .....AsnTyLeuCy 1424
      |||:|||||
3636 TGTTCCCGAGGTGTTGGCGAGGAAGAGTGTCAGCTCATCTGTTACTGC 3685
      |||:|||||
1424 sValValAlaAlaGlyLeuValAspSerIleValAspSerValValLeu 1441
      |||:|||||
3686 CATTACG .....CTAGGGGTA...TCTTTGATC 3711
      |||:|||||
1441 ysLeuSerGlyValAlaLysGluArgValLeuGlyPheLysSerLysIle 1457
      |||:|||||
3712 ATAAGGGCTTCCTTAAC ..... 3729
      |||:|||||
1458 IleLysAsnPheLeuAsnValPheArgLysAlaLysValValThrArgTh 1474
      |||:|||||
3730 .....GACA 3733
1474 rSerSerSerThrAspLeuSerGluAspGluTyPheSerCysAspGluS 1491
      |||:|||||
3734 CCATACCTCAACT ..... 3747
      |||:|||||
1491 erLysProGlyLeuArgGlyLysSerSerArgPheThrLeuSerArgLeu 1507
      |||:|||||
3747 ..... 3747
1508 LeuAspIlePheAsnPheLeuLysSerSerLysLeuValIleGluAs 1524
      |||:|||||
3748 .....GCTTAGCTCCACCGGTAGAGGGAGGATGTGTAC. 3783
      |||:|||||
1524 nAlaCysPheSerAlaTyArgGluArgIleGluArgAsnMetLysLeuTyP 1541
      |||:|||||
3784 .....GATGAGACGCTTAGGTATTACCGG... 3807
      |||:|||||
1541 hePheProLeuAsnSerSerGluGluAlaArgLeuIleArgCys 1557
      |||:|||||
3808 .....GACTTTGACTAT .....GACGAAGGTGCTGG 3833
      |||:|||||
1558 AlaGlyAspPheAspTyLeuSerSerAlaPheAspGluAspGluMe 1574
      |||:|||||
3834 TCCATCTGGGACTCAGCATGAAGCGTTCCCGGTGACGATAC ..... 3876
      |||:|||||
1574 tLeuArgGlnAlaPheGluGlnTyTySerSerAspAspGluSerValT 1591
      |||:|||||
3877 .....GATGGATCCACTTCTAGTGCTCAAGCATATGCTGTGCAAAAT 3921
      |||:|||||
1591 hrTyAspGlyLysProThrValLeuArgSerTyLeuAsnValSerArg 1607
      |||:|||||
3922 GTGGCGGACGTGGGATTAGCACCAACCGG .....GAAGT 3956
      |||:|||||
1608 ..ArgPheLeuGluThrPheCysAsnGlyProLysPhePheValLysVa 1623
      |||:|||||
3957 TACTGGTGAAGAAGACCCATTACCTCGAAGCGTGCAA...TACACTT 4003
      |||:|||||
1623 lSerAsnTyPheLysAlaLeuTySerArgLeuLeuArgValLeuProT 1640
      |||:|||||
4004 ATGTCCAGGAAGAGGTGTGCCCGCTGTGCAGCTGTGGCGGAAGACAAGGT 4053
      |||:|||||
1640 rpValAspArgAsnLeuSerAspSer ..... 1648
      |||:|||||
4054 GATCCGTGGGTTCTGGTACCGCTTACCGCTATGGCTTTTGTGAAAGTGT 4103
      |||:|||||

```

```

1649 ...ProGlyLeuLysGlyGlyAsnGluLysAlaLeuLeuAlaLysPheL 1664
      |||:|||||
4104 GAAAAAG .....GTGTCGACGATGCTTTTCCACCAACAGTCTAGTGGGG 4147
      |||:|||||
1664 euLysThrCysValIleThrAlaCysGluCysValSerGlnIleCysCys 1680
      |||:|||||
4148 AAACGGCTCGTGAGGTTGAGTGGCGGCAAGAGGTTGCTCCAGAAAGC 4197
      |||:|||||
1681 LeuArgLeuIleArgLeuCysTrpGlyThrProAlaCys ..... 1693
      |||:|||||
4198 GTCTGCGGTGAGCGCCGACACAAAGAAAGGAGAGCTGCAGATGGTAA 4247
      |||:|||||
1694 .GlyLeuValArg .....LeuPheTyIleT 1702
      |||:|||||
4248 CACACACAAACCGGG .....TCAACGAAGCGACAGGAGCCAGTAC 4291
      |||:|||||
1702 hrTySerGlyThrArgValLeuSerArgValValAlaValAlaVal 1718
      |||:|||||
4292 AGTCAGCTCTTGTGAGTTCGCCACAGCGTG ..... 4321
      |||:|||||
1719 CysProLeuLeu...ValArgAsnGluLeuAspGlyLeuSerAspGlyLe 1734
      |||:|||||
4322 .....ATATTCCAAAGTCAACCCAGTCCGAGGTACA 4352
      |||:|||||
1734 uThrAsnMetGlyValSerValPheArg ..... 1743
      |||:|||||
4353 TGCTCAGAAAGAAGTCAACAAAGAGTACCATTTGGCGACTGTTTCGGGG 4402
      |||:|||||
1744 .....ArgLeuPheValAla 1748
      |||:|||||
4403 CCACGCCAATCGTCGATGAGAAACCCGCCCAAGTGTTCAGACTCGTGT 4452
      |||:|||||
1749 LeuArgArgAla ..... 1752
      |||:|||||
4453 GTGAAGATTAATTGACAAGGCAAGCGCTCATGTGGCTGAGAAAAA 4502
      |||:|||||
1753 .....LeuS 1754
      |||:|||||
4503 ACAGGTACAAGTCGAGACGCCCAACAGAGAGTTTGACGATCAATGAAG 4552
      |||:|||||
1754 erAlaTySerAsnSerAlaLeuArgLysIlePheGluPheIlePhe 1770
      |||:|||||
4553 GCAAGGCCCGTTAAACAGCTTT .....GCATGTTTAGAACGTTTCCTGC 4596
      |||:|||||
1771 GlyAsnIleHisHisProPheAspValAlaValIleGlu ..... 1783
      |||:|||||
4597 GGTGTGCAGCTGGATGTACAAAGAGGACTATCGCCACAGGTTC 4646
      |||:|||||
1784 .....ThrAsnGluValAlaProGluProLeuS 1793
      |||:|||||
4647 AAACGATTTACCTTTGTCGATAACTTGAAGGGAGGA .....GTGCGG 4690
      |||:|||||
1793 erProGluValAspIleAspValAspCysAspPheGlySerAspSerGlu 1809
      |||:|||||
4691 TCTTTTCTCAAAGCTGGGTGAGGGGTATACCT ..... 4723
      |||:|||||
1810 SerValSerSerAspGluValAlaSerIleProArgProGlyLeuHisG 1826
      |||:|||||
4724 .....ATAATGGTGGTAGCCATGT 4742
      |||:|||||
1826 yGlySerArgArgSerSerAsnPheLeuThrSerLeuValLysValAlP 1843
      |||:|||||
4743 TTCATCAGGGTGGCTCGTGCC ..... 4765
      |||:|||||
1843 heLysLeuAlaGlyArgIleProArgLeuLeuPheArgLeuArgAsnPhe 1859
      |||:|||||
4766 .....TAGAGGATATCTTAACGCAATTAAGTACC... 4795
      |||:|||||
1860 ValAlaTyPheValGluArgArgLeuAlaSerLysArgLeuLysThrPh 1876
      |||:|||||
4796 .....CAACGGTCTTCGACCACCTGTTTAGT 4820
      |||:|||||
1876 erIleGlyLeuAlaArgLeuPheAspAsnPheSerLeuThrSerVal... 1891
      |||:|||||

```



4821	GCAGAAAGTACAAGATGGGTGGAGGCGTTACCATTTCCAGCTCATCACGAGG	4871	AGTGCTATCCATCATCAATAACCCATCTCTTGACGGTCA	4906	.....	4936	.....	4966	.....	4996	.....	5026	.....	5056	.....	5086	.....	5116	.....	5146	.....	5176	.....	5206	.....	5236	.....	5266	.....	5296	.....	5326	.....	5356	.....	5386	.....	5416	.....	5446	.....	5476	.....	5506	.....	5536	.....	5566	.....	5596	.....	5626	.....	5656	.....	5686	.....	5716	.....	5746	.....	5776	.....	5806	.....	5836	.....	5866	.....	5896	.....	5926	.....	5956	.....	5986	.....	6016	.....	6046	.....	6076	.....	6106	.....	6136	.....	6166	.....	6196	.....	6226	.....	6256	.....	6286	.....	6316	.....	6346	.....	6376	.....	6406	.....	6436	.....	6466	.....	6496	.....	6526	.....	6556	.....	6586	.....	6616	.....	6646	.....	6676	.....	6706	.....	6736	.....	6766	.....	6796	.....	6826	.....	6856	.....	6886	.....	6916	.....	6946	.....	6976	.....	7006	.....	7036	.....	7066	.....	7096	.....	7126	.....	7156	.....	7186	.....	7216	.....	7246	.....	7276	.....	7306	.....	7336	.....	7366	.....	7396	.....	7426	.....	7456	.....	7486	.....	7516	.....	7546	.....	7576	.....	7606	.....	7636	.....	7666	.....	7696	.....	7726	.....	7756	.....	7786	.....	7816	.....	7846	.....	7876	.....	7906	.....	7936	.....	7966	.....	7996	.....	8026	.....	8056	.....	8086	.....	8116	.....	8146	.....	8176	.....	8206	.....	8236	.....	8266	.....	8296	.....	8326	.....	8356	.....	8386	.....	8416	.....	8446	.....	8476	.....	8506	.....	8536	.....	8566	.....	8596	.....	8626	.....	8656	.....	8686	.....	8716	.....	8746	.....	8776	.....	8806	.....	8836	.....	8866	.....	8896	.....	8926	.....	8956	.....	8986	.....	9016	.....	9046	.....	9076	.....	9106	.....	9136	.....	9166	.....	9196	.....	9226	.....	9256	.....	9286	.....	9316	.....	9346	.....	9376	.....	9406	.....	9436	.....	9466	.....	9496	.....	9526	.....	9556	.....	9586	.....	9616	.....	9646	.....	9676	.....	9706	.....	9736	.....	9766	.....	9796	.....	9826	.....	9856	.....	9886	.....	9916	.....	9946	.....	9976	.....	10006	.....	10036	.....	10066	.....	10096	.....	10126	.....	10156	.....	10186	.....	10216	.....	10246	.....	10276	.....	10306	.....	10336	.....	10366	.....	10396	.....	10426	.....	10456	.....	10486	.....	10516	.....	10546	.....	10576	.....	10606	.....	10636	.....	10666	.....	10696	.....	10726	.....	10756	.....	10786	.....	10816	.....	10846	.....	10876	.....	10906	.....	10936	.....	10966	.....	10996	.....	11026	.....	11056	.....	11086	.....	11116	.....	11146	.....	11176	.....	11206	.....	11236	.....	11266	.....	11296	.....	11326	.....	11356	.....	11386	.....	11416	.....	11446	.....	11476	.....	11506	.....	11536	.....	11566	.....	11596	.....	11626	.....	11656	.....	11686	.....	11716	.....	11746	.....	11776	.....	11806	.....	11836	.....	11866	.....	11896	.....	11926	.....	11956	.....	11986	.....	12016	.....	12046	.....	12076	.....	12106	.....	12136	.....	12166	.....	12196	.....	12226	.....	12256	.....	12286	.....	12316	.....	12346	.....	12376	.....	12406	.....	12436	.....	12466	.....	12496	.....	12526	.....	12556	.....	12586	.....	12616	.....	12646	.....	12676	.....	12706	.....	12736</
------	--	------	---	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	---------

5464 .....AATATTGCGGTCACTACCGCGGAAGGTGTTTCGAAAGTA 5504  
2162 ysAspGluAspMetPheValLeuArgAlaGlyGlnGlyValValSerGly 2178  
5503 CCTAATCAAGTTTATGATCACTACCGGGCTTTACGTTTACAAAGTCGG 5552  
2179 ArgAsn.....SerArgLeuProLeuLysHisPhe.....LysGI 2190  
5553 CACAGATCTCATTTTCATCAACAAGACGGGTGCGTGTGAGACACC 5602  
2190 yHisGluPheCysPheArgSer.....GlyGlyLeu.....V 2201  
5603 TACCGTACCTACTCATAGCTGAAAGATATCTTTACCAAGGGCAAGAT 5652  
2201 aIProtyr..... 2203  
5653 GTCGACGCGGTGGTAGCTTGGCGCACAACTGTTCTGTATGCCGACGATAT 5702  
2204 .....AspGlyThrSerArgValAspThrII 2212  
5703 ACTGGTTTTCCAGATGCCATTAAATTCATAGGTGCATGAAGATCGCTC 5752  
2212 e.....PheHisThrGlnThrAsnPheValSerAla...AsnAlaLeuL 2226  
5753 GATGGCGCATGTGGCGGAATCGTTTAAGTCTCTCGAATATAAG..... 5796  
2226 euSerGlyTyrLeu....SerTyrArgThrPheThrPheThrAsnLeu 2240  
5797 .....TGCTATANTGCTCCCCACGGTGGCGGTATAGACGAC 5831  
2241 SerAlaAsnValLeuLeuTyrGluAlaProGlyGlyGlyThrTh 2257  
5832 GACGTTAGTGGAGAAATTCGTTAAG.....TCACCAATAGCACACGCCA 5875  
2257 rThrLeuIleLysValPheCysGlyuThrPheSerLysValAsnSerLeuI 2274  
5876 CCATTACGGCTAATGTGGGAAGTTCTTGAGGACATA..... 5910  
2274 leuThrAlaAsnLysSerArgGluGluIleLeuAlaLysValAsn 2290  
5911 AATATGGCGGTGAACAGAGATCCGAATTTGGAAGTCTCAACAGTGC 5960  
2291 ArgIleValLeuAspGluGlyAspThrProLeuGlnThrArgAspArgII 2307  
5961 TACCACGTTAACTCCAGGGTGGTAACTTTATCGTCAGGGGAATGATA 6010  
2307 leuThrIleAspSerTyrLeuMetAsn.....AsnArgGlyLeuThrC 2322  
6011 AAAGGGTTTGTG...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTA 6057  
2322 ysLysValLeuTyrLeuAspGluCysPheMetValHisAlaGlyAlaAla 2338  
6058 CAATAGGCGCTTCGCACCGCGCGTGCAGAGCGCTCTTTTGGAGA 6107  
2339 vaIAlaCysIleGluPheThrLysCysAspSerAlaIleLeuPheGlyAs 2355  
6108 CATAATACGATACCATTCATAAACACAGGCAAGGTGTTAGGATCGATT 6157  
2355 pSerArgGlnIleHisTyrIleAspArgAsnGlu.....LeuAspT 2369  
6158 GTGCTGTT.....TTGTGTCCAAAGGAAGGAAGCGTT 6189  
2369 hrAlaValLeuSerAspLeuAsnArgPheVal...AspAspGluSerArg 2384  
6190 GTATACACTTCTTAANTGTACAGGTGTCGGTTAGATGTTTGCCTACTGTGT 6239  
2385 valTyrGlyGluValSerTyrArgCysProThrAspValCysAlaTrpLe 2401  
6240 GTCCCTCAATGACCGTTAAGGGGACGGAAGAGTGTACCTCGAAGAGTCG 6289  
2401 uSerThrPhe.....TyrProLysThrValA 2410  
6290 TT.....AGCGGTAGGACAACACAGTAGTAGAATCG 6321

```

::      :::::
2410 laThrThrAsnLeuValSerAlaGlyGlnSer.....Ser 2421
6322 CTGTCCTCAAAAGCCCACTTGAACCACTGATGACGTAGCTGAATAAACGC 6371
:::      :::::
2422 MetGlnValArgGluIleGluSerValAspValGluTyrSerSerG1 2438
6372 TGACCTGTACTTGTGCATCACCACCTTGGAGAACTCGGATATG..... 6414
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2438 uPheValTyrLeuThrMetLeuGlnSerGluLysLysAspLeuLeuLysS 2455
6415 .....AAGAGTCTGTGAAGAAAGAAAGAAAGAAACACCACTGATG 6456
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2455 orPheGlyLysArgSerArgSerValGluLysProThr...ValLeu 2470
6457 ACAGTGCATGAAGCACAGGAAACATTCAGTGATGGTATCTTTAG 6506
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2471 ThrValHisGluAlaGlnGlyGluThrTyrArgLysValAsnLeuValAr 2487
6507 GACGAAGAAGCCGATGACTCCCTATTCTCACTAAACAACCGCATATCTTG 6556
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2487 gThrLysPheGlnGluAspAspPheArgSerGluAsnHisIleThrV 2504
6557 TTGGTTTTCGAGACACACACCTCACTGGTTTATCGCGCTCTGAGCTCA 6606
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2504 aIalaLeuSerArgHisValGluSerLeuThrTyrSerValLeuSerSer 2520
6607 AAGTTGGACGATAAGTCGGCCACATATATTAGCGACGCG 6645
|||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2521 LysArgAspAspAlaIleAlaGlnAlaIleValLysAla 2533
```

seq\_name: sp\_virus:089906

```

seq_documentation_block:
ID 089906 PRELIMINARY; PRT; 3070 AA.
AC 089906;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE FUSION PROTEIN.
OS Beet yellows virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID-12161;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-CALIFORNIAN;
RX MEDLINE-98285746; PubMed-9621048;
RA Permyslov V.V., Hagiwara Y., Dolja V.V.;
RT "Genes required for replication of the 15.5-kilobase RNA genome of a
RL plant closterovirus.";
J. Virol. 72:5870-5876(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-CALIFORNIAN;
RA Dolja V., Permyslov V., Hagiwara Y.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF056375; AAC25115.1;
DR INTERPRO: IPR00606;
DR INTERPRO: IPR001788;
DR INTERPRO: IPR002588;
DR PFAM: PF00978; RNA_dep_RNAPol2; 1.
DR PFAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
FT CHAIN 589 589 2042 PAPAIN-LIKE LEADER PEPTIDE.
FT CHAIN 2626 3070 PUTATIVE METHYLTRANSFERASE AND RNA
FT CHAIN 3070 3070 HELICASE.
SQ SEQUENCE 3070 AA; 345803 MW; 6FC307093F1DF36E CRC64;
```

alignment\_scores: Quality: 744.50 Length: 2220

```

Ratio: 0.731 Gaps: 103
Percent Similarity: 45.856 Percent Identity: 22.252
alignment_block:
US-09-301-906-3 x 089906 ..
Align seg 1/1 to: 089906 from: 1 to: 3070
1222 GAGGTCTATTTCATCGGTC.....GTTTCGAGTATCACCANTAG 1259
|||||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
596 GluIleThrSerSerLysLysHisLysAlaIleGluSerValTyrGluAr 612
1260 ACTGTTATTAAAGGACCAATCGGCATTTGTTTCCCATTTTGGACACGAAAT 1309
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
612 gValSerIleHisLysAspAsnLeuLeuAlaArgSerValGluLysAspL 629
1310 TGTGGGATATGTTTCTCAAGGGAGGCAATGATTCGGCGGAAAAACCTCA 1359
||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
629 euIleAspPheLysAspGluIleLysSerLeuSerLysGluLysArgSer 645
1360 CATAGGTGCGATGTGTTCTGAAGCGCGCGGAAAGGAGAACTGAGGCA 1409
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
646 ValThrValProPheTyrMetGlyGluAlaValGlnSerGlyLeuThrAr 662
1410 ACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCCAGAGTAGTC 1459
:  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
662 gAlaTyrProGlnPheAsnLeuSerPheThrHisSerValTyrSerAspH 679
1460 ACCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGG 1509
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
679 IsProAlaAlaAlaGlySerArgLeuLeuGluAsnGluThrLeuAlaSer 695
1510 AGGTGTGTAATGTGCTTCTTCGATATTGGGGGAGCTTCACGTATCA 1559
|||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||
696 MetAlaLysSer...SerPheSerAspIleGlyGlyCysProLeuPheH1 711
1560 TGTCAAAGCTGCCCATGTGAACATGTCATGTCATCAATCCAGTCCTAGACG 1609
|||||  |||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
711 sIleLysArgGlySerThrAspTyrHisValCysArgProIleTyrAspM 728
1610 TTAAGATGTGAAGGGGAGAATCAATGAGATCCTCTTTTCCACAGCT 1659
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
728 etLysAspAlaGlnArgValSerArg.....GluLeuGlnAlaArg 742
1660 GGGGAGATTCGTAGTCAGTCACCTTCTAACTGAAGCGGCTTCAAA 1709
|||||  |||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
743 GlyLeuValGluAsnLeuSerArgGlnLeuValGluAlaGlnAlaAr 759
1710 GTCTGTGTTTACTAGTCGAAATCGCAGAACTCGGATTCAGAGCG 1759
:  |||||  |||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
759 g...ValSerValCysProHisThrLeuGlyAsnCysAsnValLysSerA 775
1760 ATGCGGGTTTTATGTTGATGTAGGATATATCCCGCAGCAGGAGTAGCA 1809
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
775 spValIleIleMetValGlnValTyrAspAlaSerLeuAsnGluIleAla 791
1810 GAGCGTATGGATAAGAAGGTCGCGTGGTTTTCGACATAGCTCTTATGTT 1859
|||||  |||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
792 SerAlaMetValLeuLysGluSerLysValAlaTyrLeuThrMetValTh 808
1860 CCCCCTGGAGTTGTTGTCAGGTAACGGTGAAGTTTACTTTGGAAGAACTC. 1908
|||||  |||||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
808 rProGlyGluLeuLeuAspGluArgGluAlaPheAlaIleAspAlaLeuG 825
1909 .....GATACGTTGGTGAAGAGGAGGAGTGATTACCTGGCCTACATGTT 1953
|||||  |||||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
825 lYcysAspValValValAspThrArgAspMetValGlnTyrLysPhe 841
1954 GGTACGTGTGGTGAGATGATGAACATCTCTTCTTAAGTAAGCGGGTT 2003
|||||  |||  |||  |||  |||  |||||  |||||  |||||  |||||  |||||
842 Gly.....SerSerCysTyrCysHisLysLeuSerAsnIleLysAsnI1 856
2004 T.....TTCACCTTTTCTTATGTACGCACTTCGTCGCGGA 2038
```





```

1860 alalaTyrPheVal..... 1864
|||||...
5032 GTAACCTCCATCCAGCAAGGCCATCAGT.....TTGACGTT 5069
||| |||...
1865 .....GluArgAlaGluAlaSerLysArgLeuLysThrPh 1876
5070 CAGGGCAACTCGCGGCTCTTT.....GGTGTAGGCA 5101
| : : : ||| : |||
1876 eLleGlyLeuAlaArgLeuPheAsnPheSerLeuThrSerValVal 1893
5102 GGATGTTGAGTTAGCGCGGCTGTCGGATGAG..... 5136
: : ||||| : : ||| : :
1893 yLeuLeuGlnGluTyrAspSerValLeuAsnAlaPheLleAspValGlu 1909
5137 .....AAGTCACAGGTGTTCCAAACCACCAACACAGCCCAAGG 5177
: : ||| ||| : :
1910 LeuValLeuLeuAsnSerGlySerValAsnValLeuProLeuValSerTr 1926
5178 TGTACACAGAACATCACACCAAAATCGGGGGCAAGGCTCTATCTAG.. 5226
: : : : : ||| : : ||| : :
1926 pValArgGlySerLeuThr.....LysLeuAlaGluV 1937
5227 .....GGAAGTGT.....AGGGAAGTCAAGGGGAGG.....TCGACA 5259
||||| : : : : |||||
1937 alLeValGlySerGlyPheAlaSerPheLeuGlyArgMetCysCysArg 1953
5260 TACTCGATATGGTCGCAACAAGATTACGTTAGGAAGTGTGAGTCGCTCAG 5309
||| ||||| : : : : :
1954 ValSerAspTrpCysSerSerSerSerAsnAlaGlyCysAsnPheMet... 1969
5310 GGCTGATATCCAGTATG..... 5328
: : : |||||
1970 .....SerProValArgThrLysGlyLysPheValProSerSerS 1984
5329 .....GCTCTTGAACCTGAC 5343
||||| ||||| |||
1984 erGlySerThrAlaSerMetTyrGluArgLeuGluAlaLeuGluSerAsp 2000
5344 TACACCCCAATGACATTCAAGTGGTTAAACCGGGACCTCTGAAGATGC 5393
: : : : : ||| : : : : :
2001 IleArgGluHisValLeuSerThrCysArgValGlySerAspGluGlu 2017
5394 CGTC.....GTGGAGTACTTGAAGTATC 5416
||||| : : : : :
2017 uGluArgProLysGluValThrGluProGlyIleGluHisThrSerGluA 2034
5417 TGCTATAGCATTTAGAGGACATACA.....GGCGTTG 5451
: : : : ||| ||| : : : : ||| : :
2034 spValValProLle.ArgSerHisSerGlnProLeuSerGlyGlyGly 2050
5452 CTTATGGCTAGAA.....ATATTGCCGT 5474
| : : |||
2050 sSerTyrSerGluAspArgGluAsnGluArgAlaAsnLeuLeuPro.. 2066
5475 CACTACCGCGGAAGGTCTCGAACTACTATCAAGTTTATGAATCAC 5524
: : : : : ||| : : : : :
2067 .....HisValSerLysIleValSerGlu 2074
5525 TACCGGCTTTCAGCTTACAGTCGGGCACAG.....AT 5559
||||| : : : ||| : : :
2075 ArgArgGlyLeuGluThrAlaArgAsnLysArgThrLeuArgGlyVa 2091
5560 CTCATTTTTCATTCAACACAAGCGCTTGGCTGTGAGAGACCT..... 5603
||| ||| : : : : : ||| : : :
2091 lSerGluPheLeu.AsnAlaIleAsnThrSerAsnGluGlnProArgPro 2107
5604 .....ACCGTACGTACTCATGCTGAAA 5626
||||| : : : ||| : :
2108 lleIleValAspHisSerProGluSerArgAlaLeuThrAsnSerValAr 2124
5627 AAGGTATCTTTACCAAGGCAA.....AGATGTCA.. 5657
: : : : ||| : :

```

```

2124 gGluPheTyrTyrLeuGlnGluLeuAlaLeuPheGluLeuSerCysLysL 2141
5658 .....CGCGTGGTAGCTTTGGCGCACAATCT 5684
: : : : ||| : : : : ||| : :
2141 euArgGluTyrTyrAspGlnLeuLysValAlaAsnPheAsnArgGlnGlu 2157
5685 GTTCGTATCGCAC.....GATATACTGGTT..... 5709
: : : ||||| : : : : : |||
2158 Cys..LeuCysAspLysAspGluAspMetPheValLeuArgAlaGlyGlnG 2174
5709 ..... 5709
2174 lyValValSerGlyArgAsnSerArgLeuProLeuLysHisPheLysAsp 2190
5709 ..... 5709
2191 HisGluPheCysPheArgSerGlyGlyLeuValProTyrAspGlyThrSe 2207
5710 .....TTCACGATGCCATTAAATTTGATAGGTGCTACTGA 5743
||||| : : : : : ||| : :
2207 rArgValAspThrIlePheHisThrGlnThrAsnPheValSerAla...A 2223
5744 AACTCCCTCGATCGGCATCGTGGCGCAATCGTTAAGTCTCTTCCAATAT 5793
: : : : ||| : : : : ||| : :
2223 snAlaLeuLeuSerGlyTyrLeu.....SerTyrArgThrPheThrPhe 2237
5794 AAG.....TGCTATAATGCTCCCGCAGGTGGCGG 5822
||||| : : : ||||| |||||
2238 ThrAsnLeuSerAlaAsnValLeuLeuTyrGluAlaProProGlyGlyG 2254
5823 TAAGACAGACGCTTACTGACGAAATTCGTTAAG.....TCACCCCAATA 5866
||||| ||||| : : ||| : :
2254 yLysThrThrThrLeuIleLysValPheCysGluThrPheSerLysValA 2271
5867 GCACGCCACCATTCAGGCTAATGTGGGAAGTCTCGAGGACATA..... 5910
: : : : ||||| : : ||| : :
2271 snSerLeuIleLeuThrAlaAsnLysSerSerArgGluGluIleLeuAla 2287
5911 .....AATATGGCGGTGAAGAAGACAGATCCCAATTTGGAAGTCT 5951
: : : : : ||| : : ||| : :
2288 LysValAsnArgIleValLeuAspGluGlyAspThrProLeuGlnThrAr 2304
5952 CACACGTGTACCACAGTTAACTCCAGGGTGGTAAACTTTTATCGTCAGGG 6001
: : : ||| : : : ||| : :
2304 gAspArgIleLeuThrIleAspSerTyrLeuMetAsn.....AsnArgG 2319
6002 GAATGTATAAAGGTTTG...GTGATCAGGTGCACATGATGCATCAA 6048
: : : : ||||| : : ||||| : :
2319 lyLeuThrCysLysValLeuTyrLeuAspGluCysPheMetValHisAla 2335
6049 GGCTTACTACAACTAGCGCTTTCGCAACCGCGCTCGGAAGGCTCTT 6098
||| : : : ||| : : : : :
2336 GlyAlaAlaValAlaCysIleGluPheThrLysCysAspSerAlaIleLe 2352
6099 TTTTGGAGACATAAATCAGATACCATTCAATAACAGGGAGAGGTGTTTA 6148
||||| : : : ||||| : : : |||||
2352 uPheGlyAspSerArgGlnIleHisTyrIleAspArgAsnGlu..... 2366
6149 GGATGGATTCTGCTGTT.....TTTGTTCCAAAGAAG 6180
: : ||| |||||
2367 ..LeuAspThrAlaValLeuSerAspLeuAsnArgPheVal...AspAsp 2381
6181 GAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCGGTTAGATGTTTG 6230
||||| ||||| : : ||||| |||||
2382 GluSerArgValTyrGlyGluValSerTyrArgCysProTrpAspValCy 2398
6231 CTACTTGTGTTCTCAATGACCGTAAAGGGACGGAAAGTGTACCCCTG 6280
| : ||||| : :
2398 saLatrpLeuSerThrPhe.....TyrProL 2407
5281 AAAGGTCTGTT.....ACCGTAAAGCAACACGATGA 6312
: : ||| : :
2407 yThrValAlaThrThrAsnLeuValSerAlaGlyGlnSer..... 2420

```

```
6313 CTAAGATCGCTGCTCAAAAGGCAATTTGAACCACTGATGACGTAGCTCA 6362
||||:||||: ||| ||| :||:||||:||||:
2421 .....SerMetGlnValArgGluLeuGluSerValAspValGluTyr 2435
6363 AATAACCGTCAGCTGCTACTTGTGTCATGACCCAGCTGGAGAACTCGGATA 6412
::: ||||| ||| ||| |||||:||||:
2435 rSerSerGluPheValTyrLeuThrMetLeuGlnSerGluLysLysAspL 2452
6413 TG.....AAGAGTGGTTGAAGGGAAGAAAGAAACA 6447
:: ||||| ||| ||| ||| |||
2452 euLeuLysSerPheGlyLysArgSerArgSerSerValGluLysProThr 2468
6448 CCAGTGATGACAGTCATCAACGACACGAGAAACATTCAGTGATGGT 6497
||||:||||:||||:||||:||||:||||:
2469 ...ValLeuThrValHisGluAlaGlnGlyGluThrTyrArgLysValAs 2484
6498 ATTGTTTAGGACGAGAAGCGGATGCTCCCTATTCTACTAAACACCGC 6547
||| ||||| ||| ||| ||| |||
2484 nLeuValArgThrLysPheGlnGluAspPropheArgSerGluAsnH 2501
6548 ATATACTTGTGTTGTCGACACACACGCTCACTGTTTATCCGCT 6597
||||| |||||:||||:||||:||||:||||:
2501 tsileThrValAlaLeuSerArgHisValGluSerLeuThrTyrSerVal 2517
6598 CTGACGCTAAAGTTCGACCATTAAGCTCGGCACATATATTAGCGACGC 6645
||||| ||||| ||||| |||||
2518 LeuSerSerLysArgAspAspAlaIleGlnAlaIleValLysAla 2533

seq_name: sp_virus:Q08534

seq_documentation_block:
ID Q08534 PRELIMINARY; PRT; 3074 AA.
AC Q08534;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (CONTAINS: 66 KDA
DE PROTEIN).
OS Sugar beet yellow virus (SBYV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=31714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UKRAINIAN (BYV-U).
RX MEDLINE=94082464; Pubmed=8259666;
RA Agranovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
RA Lunina N.A., Atabekov J.G.;
RT "Beet yellows closterovirus: complete genome structure and
RL virology 198;311-324(1994).
CC -!- FUNCTION: PROBABLE POLYMERASE.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
CC THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
CC EMBL; X73476; CAA51871.1; -.
DR MEROPS: C42.001; -.
DR INTERPRO: IPR000606; -.
DR INTERPRO: IPR001788; -.
DR INTERPRO: IPR002588; -.
DR PRAM; PF00978; RNA_dep_RNAPol2; 1.
DR PRAM; PF01443; Viral_helicase; 2.
DR PRAM; PF01660; Vmethyltrans; 1.
KW RNA-directed RNA polymerase.
FT CHAIN 1 3074
FT CHAIN 1 588
FT MUTAGEN 509 509 C-X: NOT PROCESSED INTO ITS SUBUNITS.
FT MUTAGEN 517 517 C-X: REDUCED PROCESSING INTO SUBUNITS.
FT MUTAGEN 518 518 C-X: REDUCED PROCESSING INTO SUBUNITS.
FT MUTAGEN 556 556 H-X: NO EFFECT.
FT MUTAGEN 569 569 H-X: NOT PROCESSED INTO ITS SUBUNITS.
FT MUTAGEN 588 588 G-X: NOT PROCESSED INTO ITS SUBUNITS.
SQ SEQUENCE 3074 AA; 346028 MW; D75B9564658BEB0C CRC64;
```

```
alignment_scores:
Quality: 740.50 Length: 2504
Ratio: 0.647 Gaps: 120
Percent Similarity: 45.687 Percent Identity: 22.764

alignment_block:
US-09-301-906-3 x Q08534
..
Align seg 1/1 to: Q08534 from: 1 to: 3074
409 CGGTATGCGCCCGTCAGCGCGAGCGAGCCGCTAAACCGGCTGTTAAAGCAACA 458
||| ||| ||||| ||||| ||||| ||||| |||||
319 ProArgLeuProLysPheArgGluValGluIleProValValLysLysHl 335
||| ||| ||||| ||||| ||||| ||||| |||||
459 TGAG.....CAGAAGCCCAACGGCAAGCGTCACATTGGG 493
||| ||| ||||| |||
335 sAlaValProAlaValValSerLysProValArg.....T 347
494 CTGTTTAAAGCCACAGCT...GTTGGCGTCCAGCTACCTCTTCTTAAAAA 540
||| ||| ||||| ||||| ||||| ||||| |||||
347 hrPheArgProValAlaThrThrGlyAlaGluTyrValAsnAlaArgAsn 363
||| ||| ||||| ||||| ||||| ||||| |||||
541 CAGGAAGCACTGGAGCGCGCAATCAGTCCACACACAGCTCGTTGGAGCA 590
||| ||| ||||| ||||| ||||| ||||| |||||
364 GlnCysSerArgArgProArg.....AsnHisProIleLeuAr 376
591 GAAGCGCGCTTGACGTTTGGCCTT.....TTCCTCA 622
||| ||| ||||| ||||| ||||| ||||| |||||
376 gSerAlaSerTyrThrPheGlyPheLysLysMetProLeuGlnArgPheM 393
623 GTAAAGGTGGGGTGATGAGACGCGGTGTCATCTTGGGGAAGGAAA 672
||| ||| ||||| ||||| ||||| ||||| |||||
393 etLys.....GluLysLysGluTyrTyrValLysArgSerLys 405
673 TTGTTTAAAC.....AGGCGCTTAATGTTCTATTGAT..... 705
||| ||| ||||| ||||| ||||| ||||| |||||
406 ValValSerSerCysSerValThrLysSerProLeuGluAlaLeuAlase 422
706 ....GTAAGAAGACAGCTGCTGTTTGGGCT.....AAAA 733
||| ||| ||||| ||||| ||||| ||||| |||||
422 rIleLeuLysAsnLeuProGlnTyrSerTyrAsnSerGluArgLeuLysP 439
734 TCTGGGATGAACCTCTCGTAGGAGGGGTATTTTACGTCAAGATAGA 783
||| ||| ||||| ||||| ||||| ||||| |||||
439 heryrAsp.....HisPheIleGlyAspAspPhe 448
784 GCTGTAAATTCCTCCCTATTGCGGGGTAGGCTACGATC..... 825
||| ||| ||||| ||||| ||||| ||||| |||||
449 GluIleGluValHisProLeuArgGlyGlyLysLeuSerValLeuL 465
826 .....GAGGACTTCATGTAATACAGCC.....CCAGGT 856
||| ||| ||||| ||||| ||||| ||||| |||||
465 eLeuProLysGlyGluAlaTyrCysValValThrAlaAlaThrProGlnT 482
857 GTGATGTTGCCCTTG.....CCGCGCAT...GAG 882
||| ||| ||||| ||||| ||||| ||||| |||||
482 yrHisAlaAlaLeuThrIleAlaArgGlyAspArgProArgValGlyGlu 498
883 TTGTGAGTATCGCGGAAAGGGCGTTTGTATGCACCAACCAAGGGTGGT 932
||| ||| ||||| ||||| ||||| ||||| |||||
499 LeuLeuGlnTyrArg.....ProGlyGluGlyLeuCy 509
933 TTGTTTAAACATGAGAGCTGAGGGGAGAAATTTACAGACGCTCGTGTGT 982
||| ||| ||||| ||||| ||||| ||||| |||||
509 sfyrLeuAlaHisAlaAlaLeuCysCysAlaLeuGlnLysArgThrPheA 526
983 TCTCATCTCTCTTTTCGATAGGTTTCTTGATGCACCTTGGCTTAGATCG 1032
||| ||| ||||| ||||| ||||| ||||| |||||
526 rgLluGluAspPhePheValGlyMetTyrProThrLysPheValPheAla 542
1033 TTAAAGGTTCATTAGGTTTGGGGGCACGAACATACATGCCATCACT 1082
```

543 LysArgLeuThrGluLysLeuGlyProSerAlaLeuLysHisProValAr 559  
 1083 CAATGAAGACGCTACCTTTGGTGGAGGCGGAGACGTC..... 1122  
 559 gGlyArgGlnValSerArgSerLeuPheHisCysAspValAlaSerAlap 576  
 1123 .....TATCTCCCAANTGTCCCAAAACCGCTATCGTC 1155  
 576 heSerSerProPheTyrSerLeuProArgPhe..... 586  
 1156 GCTGGCATAGGACACGGTTGGGAGG.....GAGAT 1187  
 587 .....lleGlyGlyValGluGluAlaProGluIl 597  
 1188 CTTGGCCCTCGCTGCCAATGCCCTTAATCAAGAGAGGCTTATTCATCGG 1237  
 597 eThrSerSerLeuLysHis.....LysA 605  
 1238 TCCTTTCCGATATCACCAATAGACTGGTATTAAAGGACCNAATCGGCATTG 1287  
 605 laileGluSerValTyrGluArgValSerlleHisLysAspAsnLeuLeu 621  
 1288 CTTTCCCATTTGCACACGAAATGTGCGATATGTTTCTCAAAGGACGC 1337  
 622 AlaArgSerValGluLysLeuPheHisPheLysAspGluLeuLysSe 638  
 1338 AATGATTCGGGAAAAACCTCACATAGGTGCGATGTGTTCTCAAGCCCGC 1387  
 638 rLeuSerLysGluLysArgSerValThrValProPheTyrMetGlyGluA 655  
 1388 GGAAGAGGAGACCTGAGGAACTCTTCCACAGGCTTTCGATACAGTTC 1437  
 655 laValGlnSerGlyLeuThrArgAlaTyrProGlnPheAsnLeuSerPhe 671  
 1438 TCCGACTCGGTACAGGTAGTACCATTCGTAATGCCATCGGAGCTG 1487  
 672 ThrHisSerValTyrSerAspHisProAlaAlaAlaGlySerArgLeuLe 688  
 1488 TTTCAATGAATCTTTTCCAGGAGGTGGTAAATGTGCTCTCTTCGATA 1537  
 688 uGluAsnGluThrLeuAlaSerMetAlaLysSer...SerPheSerAspI 704  
 1538 TTGGGGGGAGCTTCAGGTATCATGTCAAGCTGGCCATGTGAACTGTCTAT 1587  
 704 leglyGlyCysProLeuPheHisIleLysArgGlySerThrAspTyrHis 720  
 1588 GTATGCAATCCAGCTCTAGACGTTAAAGATGTGAAGCGGAGATCAATGA 1637  
 721 ValCysArgProIleTyrAspMetLysAspAlaGlnArgValSerAr 737  
 1638 GATPCTCTTTCTTCCACAGCTGGGGAGATTCGTACGTGCTCAGTGACC 1687  
 737 g.....GluLeuGlnAlaArgGlyLeuValGluAsnLeuSerArgGluG 752  
 1688 TTCTACTGAAGCGGCTTCAAGCTGTGCTCTTACTGTAGTCAGAAATCG 1737  
 752 lnLeuValGluAlaGlnAlaArg...ValSerValCysProHisThrLeu 767  
 1738 CAGAAGCTGGATTCTAGACCGATCGGGTTTATGTTGATGCTGTACGA 1787  
 768 GlyAsnCysAsnValLysSerAspValIleMetValGlnValTyrAs 784  
 1788 TATATCCCGCAGGAGGTAGCAGAGCTATGATAAGAGGGTGCCTGG 1837  
 784 pAlaSerLeuAsnGluIleAlaSerAlaMetValLeuLysGluSerLysV 801  
 1838 TTTTGCACATAGCTCTATGTTCCCGCTGGAGTTGTTGACGGAACGCT 1887  
 801 aAlaTyrLeuThrMetValThrProGlyGluLeuLeuAspGluArgGlu 817  
 1888 GAAGTTTACTTGAAGAACTC.....GATACGTTGTTGAGAGGGAAGG 1931

818 AlaPheAlaIleAspAlaLeuGlyCysAspValValValAspThrArgAr 834  
 1932 TGATTACTGCGCTACAAATGTTGGTCAGTGTGGTCAGATGTATGAAACATT 1981  
 834 gAspMetValGlnTyrLysPheGly.....SerSerCysTyrCysHisL 849  
 1982 CCTTCTCTAACGTAAGCGGGTTT.....TTCACCTTTTCT 2016  
 849 ysLeuSerAsnIleLysSerIleMetLeuThrProAlaPheThrPhe... 864  
 2017 TATGTCACGCACCTTCGCGGGAACGTGTTAAAGCTAGAGTATGAGGGATA 2066  
 865 .....SerGlyAsnLeuPheSerValGluMetTyrGluAs 876  
 2067 CCCTGTGTGTACCATCATCTCACTATGTAGGGCTCAGAAGTCACCTG 2116  
 876 nArgMetGlyValAsnTyrTyrLysIleThrArgSerAlaTyrSerProG 893  
 2117 GA.....ACTGAGGTT 2127  
 893 luIleArgGlyValLysThrLeuArgTyrArgArgAlaCysThrGluVal 909  
 2128 ACGTATAGGTGCTGTCGCTCGTTCGTCGGCAATCGCTGGTGTTCAT 2177  
 910 ValGlnValLysLeu...ProArgPheAspLysThrLeuLysThrPheLe 925  
 2178 ACCTGTGTAGCTGGTTCATGTGTCTTAAAGACAATAGTCTCTCGATT 2227  
 925 uSer.....GlyTyrAspTyrIleTyrLeuAspA 935  
 2228 CGACATTTGTCGACAGGATCTATTCACGCGCTCAACACTATAGGACA 2277  
 935 laLysPheValSerArgValPheAspTyrValValSerAsnCysSerVal 951  
 2278 TTCGAGAATAAGCTTTGAGTATGCCGTTGGCGGCTCAGTCGCAAAA 2327  
 952 ValAsnSerLysThrPheGluTrpValTrpSerTyrIleLysSerSerLy 968  
 2328 GACCCATGTCTTACAGGAGTCGCGTTGTCCACAGCAAGCTTGATATTT 2377  
 968 sSerArgValValIleSerGlyLysValIleHisArgAspValHisIleA 985  
 2378 CTCTCTGATGATGTGGGGTTTGTCTGCTGCTGTTATGGCTCAGCGCAT 2427  
 985 spLeuLysHisSerGluCysPheAlaAlaValMetLeuAlaValGlyVal 1001  
 2428 AAGGATAGG.....GCGAAGAGTATTCCCTCCTCTATAA 2459  
 1002 ArgSerArgThrThrThrGluPheLeuAlaLysAsnLeuAsnTyrTyrTh 1018  
 2460 CTTTATAAAAGCCAGTCAGGGAGTCGCGCGGGGTCTTCAAGCTCTTC. 2508  
 1018 rGlyAspAlaSerCysPheGluThrIleArgPheLeuPheArgGluTrps 1035  
 2509 .....TTTCAGACCGTAGGCGATTGTTTTCGAACGCGAGTCTCC 2547  
 1035 erArgArgAlaTyrAlaGluIleAsnArgSerPheArgLysLeuMetLys 1051  
 2548 GTCTATGCTAAGCAATGGTGCACGATAAATCTC.....AACGT 2585  
 1052 SerIleLeuSerAlaGlyLeuAspTyrGluPheLeuAspLeuAspAsnSe 1068  
 2586 TTTTGGACACGCTTATGTCTATGCCAGAGCGCTTCATCCGTAAGATACCTG 2635  
 1068 rLeuGlnHisLeuLeuGluTyrSerGluValGluValArg..... 1081  
 2636 GGTCTGTGTGTGTACATTGTCACCTTCTGGAGCTTCAGACAGGTGGAG 2685  
 1082 .....ValSerIleAlaGlnAsnGlyGluValAsp..... 1091  
 2686 CTGAGGGGTCCCTTGTATATTTCCAGGACACCTTCGTTAGGAAACTGAA 2735  
 1092 .....Cy 1092

```
2736 GAATAGTCGCTTGGCGCTCTCTCTAGGCTATCTGTGAAGATTCAATTA 2785
|||||:|||||:|||||:|||||:|||||:
1092 sAsnGluGluAsnArgValLeuThrGluIleAlaGlu..... 1105
2786 AGGTCAATGAAGCAATGAGACAGAGATGGTGAACCCCTGCCAATTACT 2835
|||||:|||||:|||||:|||||:|||||:
1106 .....AlaAlaAspArgLys..... 1110
2836 GAAGATTCTGTATATGCTTCATATGCGGAACGTTTCTAAGCTCCACTG 2885
|||||:|||||:|||||:|||||:|||||:
1111 .....SerIleAlaGlnGlyLeuSerGlyAlaLeuSerValProth 1125
2886 T...ACGAGGGCAGGTCTCTTGGCGGTTGCGAAGCGAC... 2922
|||||:|||||:|||||:|||||:|||||:
1125 rClnProArgGlyGlyLeuArgGlySerArgArgSerGlyValSerP 1142
2923 .....GTGTTTCGAGTGTTCTAAGGGTTTGGTAGCTCGTGGG 2961
|||||:|||||:|||||:|||||:|||||:
1142 heLeuTyrAsnLeuValGluGluValGlyAsnLeuPheSerValGly 1158
2962 CGTGGCAGCAAGCGCTTCTTGGCATTACGCTGTTCTTTTCCACAGGTC 3011
|||||:|||||:|||||:|||||:|||||:
1159 aspAlaValArgPheLeuValValPheLysThrPheSerAspSerP 1175
3012 ACTATTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGG 3061
|||||:|||||:|||||:|||||:|||||:
1175 oIlePheArgValValArgMetPheLeuAsp.....LeuA 1187
3062 TCGGCACAGAGAACTCTATAACTACCGGTGGGCATCTGGAGACGTCG 3111
|||||:|||||:|||||:|||||:|||||:
1187 IaGluAlaAlaSerProPheValSerValValSerLeuCysAlaTrpLeu 1203
3112 CGCGTAGCTGTGAGC.....AAGTCGTAGC 3137
|||||:|||||:|||||:|||||:|||||:
1204 ArgGluAlaValSerAlaPheSerSerTrpValAlaAspArgThrValSe 1220
3138 TGGAACGAAGAATTTTGGAGTGAAGTTTCTTAAATGACTTCACCACTT 3187
|||||:|||||:|||||:|||||:|||||:
1220 rGluSerValLysThrPheValAsnArgThrValLysArgPheLeuAsnP 1237
3188 TCGTATTCCGAATAAGTGCTTATCGGGATATTCGTGGCGTCTTTGGGT 3237
|||||:|||||:|||||:|||||:|||||:
1237 heMetSerAlaLysThrLeuThrLysLysPhePheArgPheLeuSer 1253
3238 CGGCGCCCAATTGCA..... 3252
|||||:|||||:|||||:|||||:|||||:
1254 AlaSerAlaLeuAlaLysThrValValArgLysAlaLysValIleLeuGl 1270
3253 .....TGGAAG...TATAGCGCGGAATTGCGGCTAACGCTAGAAGGT 3292
|||||:|||||:|||||:|||||:|||||:
1270 uAlaTyrTrpGluValTrpPheGluSerIleLeuSerAspSerGlyGluT 1287
3293 AGCGGGGAGTAGTTACGAACCTAAGCTCGTTAAGTTCACACCGGCC 3342
|||||:|||||:|||||:|||||:|||||:
1287 yr.....SerAlaValGluPheCysSerSerValVal..... 1297
3343 GGTGGTTTACGGGTTTAACTCTAGCAGTATCCGGTGGATCTTTAGT 3392
|||||:|||||:|||||:|||||:|||||:
1298 .....IleThrLeuLeuThrAsnSerGlyArgLeuLeu... 1308
3393 CGTGGCAAGAGGTTTTCGTGGCGGTGACCGTCACTAGG..... 3432
|||||:|||||:|||||:|||||:|||||:
1309 .....ProGlyPheSerProSerAlaIleIleThrGluValLeuLeuA 1323
3433 .....GCGACCGTAGCTAAAGCTCAAGTCCCTTAGCTTAGCTATCGTTT 3477
|||||:|||||:|||||:|||||:|||||:
1323 spLeuAlaThrLysIleSerIleGluValLeuLeuLysGlnIleSerPro 1339
3478 TCTACCTCATACCGCATTCGGC..... 3501
|||||:|||||:|||||:|||||:|||||:
1340 AlaAspSerThrAlaSerSerAlaLeuTyrArgArgValLeuSerGluIrl 1356
4032 AGCTGTGCGGAAAGACAAGGTGATCCGTCGGGTTCTGTGTACCGCTGACG 4081
```







Percent Similarity: 44.203 Percent Identity: 20.537

alignment\_block:

US-09-301-906-3 x Q66237 ..

Align seg 1/1 to: Q66237 from: 1 to: 3124

```
139 TCGGACTTCACACCCAGGTCGCGAGCTATGATTACATTGGAAGCTCAC 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 SerAsnPheAspGluLeuLeuHisAlaVal.....LeuThrArgMetSe 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 CAAAGGGGTGAAGCGTACGTTTGTCCCGCCACAGTTAA..... 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 rLysGlyThrLeuGluLeuLeuHisSerIleArgGluArgLysG 625
229 .....GGTTTGCAGGCGAGTACGCTGTTCTCAGCGCTCAGTCAGC 270
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 lLeuLeuGlyPheAlaAlaCysPheArgTrpGlySerSerHisValPhe 641
271 GCCTCGAGAGGGAT..... 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 ValLysArgGlyAsnPheThrAlaArgPheProIleSerAlaValArgTy 658
286 .....GPTAAGAAGCTTTGAGGCAAGGACCTCAACTCCG 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 rGluAsnGlnTyrAlaAlaValMetMetAspValGlyThrSerTrpValP 675
326 CAACTTCCAGCGTGTCTGATTTCGACGCTGCTATTCGAAGCTGTTCTAAT 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 rLysGlyTyrValProAspValAspValValLysArgArgSerArg 691
376 GCATTACTTCTCTACACTACACCGCGGTAGTCCGCTATGCCCGCCGTCAA 425
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 AsnArgArgPheProLeuSerHisArgAlaVal..... 702
426 GCAGCGAGCGCTAAACCGCTCTTAAGCAAGATGAGCAAGCCCAAC 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 .....AspGlnSerValProAlaLeu.....GlnSerSerProValG 715
476 GCAAGCGTCACATGGCTGTGTAAAGCAAGCTGTTGGCGTCCACGTA 525
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
715 lValAlaSerValSerSerAspLysProSer..... 726
526 CCACCTTCCTAAACCAAGCAAGCAAGCTGGAGCCAGCCCAATCAGTCCACA 575
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 .....SerGlnLysAlaSerSerProSerThrSerSerSerSe 739
576 ACAGTCGTTGGAGGAGAGCGCCCTTGACGTTTGGCTTTCTTCAGTA 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
739 rSerPheAsnSerThrArgSerSerAspTrpValAsnValGlyPheSerH 756
626 AA.....GGTGGGGTGATGAG 642
756 lSerGlnThrAsnLysValArgGluProArgTyrAlaGlyHisArgArg 772
643 ACCGACGCTGTCATCTTCGGGAAGGAAATGCTTAAAC..... 681
773 lLeGlyAsnPheThrPheProArgGlyThrValTyrAsnThrProValAs 789
682 ....AGGCGCTTAATGTTCTTCCTATTCATGTTAAAGAACACGCTTCGTTGG 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
789 pGluArgAlaTyrLysArgValLeuArgLeuArgAspThrThrAlaCysS 806
728 CTAAATCTGGGATGAAGCTCTCGTAGGAGAGGTTATTTTACGTCAA 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
806 erPheLeuArgIleLeuLeuAlaArgLeuSerGlyTyrArgProLeuSer 822
778 GATAGAGCTGTTAAATCTTC..... 798
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
823 Asp.....GluPhePheAsnLysCysValThrSerArgPheValAl 836
799 .....CCTATTGTCGGGGT..... 813
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
836 aCysIleGluProValGluSerGlyLeuValLysValHisPheArgSerA 853
814 .....AGGCTACGATCGAGGACTTCATCGTGAATACAGCCCGAGG 855
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
853 spValPheArgAlaSerPheProPheAspGlyGlnProIleHisProAla 869
856 TGTGATGTTCCCTTCCCGCCATTCAGTTCTGGAGTATGCCGAAAGGC 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
870 ThrAlaLeuThrLeuGluValValSerTrpMetValThrAl 886
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 GTTGTGATGCACCAACAAA...GGTGGTGTGTTGTTTAAACAAT..... 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 aPheGlnAsnProLeuLysAspGlyTyrCysTyrIleArgHisPheAlaG 903
945 ..... 945
903 lValSerLeuSerMetGlyArgIlePhePheArgArgAspValAspLeu 919
946 .....GAGAGGCTC... 954
920 GlyProPheProTyrValPheGluValGlnHisArgLeuGluArgLeuTy 936
955 .....AGGGAGAAATTTACAGACGTC 976
936 rGlyLysAlaAlaLeuArgTyrGlyValArgGlyGlnTyrSerAlaProA 953
977 GTTGCTTCTCATCTCTTTCGATAGTTTCTTGATGCACCTTGGCTTT 1026
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
953 rGysPheHisCysCysTyrAsn..... 960
1027 AGATCGTTAAAGTTCATTAGTTTCCGGGCAAGCAATACATACACATGCC 1076
961 .....AspSerProArgProMetAl 967
1077 ATCACTCAATGAAGAGCGTACCTTTGGTGAAGGGCGAGAGCTCTATC 1126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
967 aSerPheAsnGlyTyrHisLysMetGlyGlyGluAspAsnSerLeuIleI 984
1127 TCCCCAATGTCCAAAACCGCTATCGTCGCTGGCGATAGACACGCTTG 1176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
984 lThrAspThr.....AspArgLeuArgAla 992
1177 GGAGGGGAGATCTTGGCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGT 1226
993 ValGly.....SerAl 996
1227 CTATTCATCGTGTTCGAGTATCACCATAGACTGGTATTAAAGGAGCC 1276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 aTyrGluLysValLysArgAsnIleProAspSerLeuLeuValArgSerV 1013
1277 AATCGCATGCTTCCCATTTGGACACGAAATTTGCGGATATGTTTCT 1326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1013 alGluLysAspLeuIleGluPheAsnGlnThrLeuValAspMetHis... 1028
1327 CAAAGGACCAATGATTCGCGAAAACCTCACATAGTGCATGTGTT 1376
1029 .....ArgSerArgProGlyValValValProPheG 1039
1377 TCTGAAGCCGCGGGAAGGAGAGCTGAGGGAACCTTTCACAGAGCTTT 1426
1039 nMetSerGluAsnGlnValMetLeuThrArgAlaTyrProGluPheA 1056
1427 CGATACAGTTCTCCGACTCGGTCCAGGAGTAGTCACCCATTTCGTAATGCC 1476
1056 snIleAsnPheIleHisSerValHisSerAspHisProValAlaAlaGly 1072
1477 ATCCGAGCTGTTTCAATCGAATCTTTTCCAGGAGGTGCTGTAATGTGTG 1526
1073 SerArgAlaLeuGluAsnHisLeuValArgLysHisAlaGly...ThrAs 1088
1527 CTCTTCTGATATTGGGGAGCTTCAGGTATCATGTCAAAGCTGGCCATG 1576
1088 pTyrSerAspValGlyCysProLeuPheHisLeuArgAlaGlyHis 1105
```



```

1670 yrArgSerValValSerPheSerAlaGluValLeuGlyLysSerTyrSer 1686
      :::::::::::::: ::::: ::::: ::::: ::::: :::::
2983 GGCATTACGTCGTTCTTTCCACAGGTTCACTATTCTACACCGCGGTTT 3032
      :: :::::
1687 LysLeuCysGlnPheCysLeuGlu.....IlePheAspIleGlyLe 1700
3033 AACTGAAGATGAAGGCTTGAT.....G 3055
1700 uileGluValSerAlaArgAsnLeuGlyLeuGluThrCysValSerIleA 1717
3056 CTCTGGTGGCCACAGAGATGCTATAAATCAACCGGTGGC..... 3096
      ::::: :::::
1717 laLeuAlaIlePheAspValIleThrPheProLeuGlySerValSer 1733
3097 .....ATACGTGGAGACGTC 3110
1734 PheProValValCysAlaArgValPheGlyArgLeuAlaMetGluValG1 1750
3111 GCCCGTAGCTGCAGCAAGTCGTAGCTGGAACGAAAGATTTTGAGTG 3160
      ::::: :::::
1750 ySerAsnPheLeuPheLysLysValPheGlyLysProGluThrLeuGlyG 1767
3161 AAGTTTCCTTAAATGACTTCACCACTTTTCGTATTGCGGAATAAGTG... 3207
      :: :::::
1767 luAspThrPheArgThrValSerMetValLeuArgProLysAlaPhe 1783
3207 ..... 3207
1784 AspLysLeuSerValAspAlaLysGlyLeuValArgCysSerGlyValI1 1800
3208 .....CITTATCGGATATTCGTGGCGCTTTGGTGGCGGCCCAATTC 3250
1800 eProGlyValLeuArgIleLeuSerSerLeuPheSerGluAspAspA 1817
3251 CATGG..... 3255
1817 laThrValGlyTyrSerLysHisGluValSerSerLeuProIleValThr 1833
3256 .....AAGTATAGCGCGGAATTCGGCTAA 3281
1834 PheCysValLysArgLeuSerTyrSerLeuGluArgCysLysGluAlaCy 1850
3282 CGCTAGAGGTACCGGGAGTAGTTACGAACTCTAAGCTCGTTAAGTT 3331
1850 sValAlaPheIleValAspSerAlaLysLysIleLeuLysSerLeuThrs 1867
3332 CACAA.....GCCGCGGT 3345
1867 erGluPheAsnGluSerIleGlnAsnSerAspLeuThrSerGlyAlaLys 1883
3346 GGTTTACGCGGTTAACTCTACACAGTATCCGGTGGATCTTTAGTCGT 3395
1884 SerLeuTyrGlyHisValSerSerIleLeuTyrThrSerSerGlyValI1 1900
3396 GCCAAGAGGTTTCGTGCGCGGTGACCGTCACTAGG..... 3432
1900 eAsnLysAlaGlySerAlaValAlaThrValLysAlaValGlySerAspI 1917
3433 ..CGCACCCTAGCTAAACGTCAGTCCCTTACCGTTGCTATCG..... 3474
1917 leSerGlyArgValLysArgMetArgProArgPheSerLeuSerArgHis 1933
3475 .....TTTTCCTACCTACGCAATTCGCGGTGCAAGTATGTT 3512
1934 PheAspGluClnThrSerGluTyrTyrSerAlaSerAspCysSerGluLe 1950
3513 A.....GCCATTTGGGCACATGCTCTTC 3535
1950 uAspGluSerLeuPheSerGluThrProGlyLeuArgGlyAsnAlaArgC 1967
3536 CACGGCACTTAATGTTTTCCTTTAGGGACATTCGTTGGGCGGAGG 3585
      :::::

```

```

1967 ySArgAspLeuPhe.....ArgIleValSerArgArg 1977
3586 GCTAGCGCGGAATACTTGGAAAGTTTGGAGGCTTCTCCAATAATTTGGTGGC 3635
      :::::
1978 AlaValAla.....CysCy 1982
3636 TGTTCGCGAGGTTCTTGGCGGAGGAAGAGTGCAGCTCATTTGTTACTGC 3685
1982 s.....GluTyrValThrIleLeuLeuLysA 1991
3686 CTATTACGCTAGGGTATCTTTGATCATAAAGGGCTTGTCTTAACGACACC 3735
      :::::
1991 rgLeuAlaSerArgAlaGlyLysTyrSerTyrSerTyrValMetAspThr 2007
3736 ATACCTCAACTGCTTACGTCGCCACCGGTAGAGGGGAGGAATGTGTACGA 3785
      :::::
2008 LeuValArgLeu...TyrVal.....AlaThrAs 2016
3786 TGACACCGCTTAGGTATTACCGGACCTTTGACTATGAC..... 3822
      :::::
2016 pAspValValArgArgCysArgSerPheThrPheAspCysThrAspLysA 2033
3823 .....GAAGGT 3828
2033 laSerAsnPheValThrGlyPheLeuSerTrpArgSerLeuArgSerGly 2049
3829 GCTGGTCCATCTCGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGA 3878
      :::::
2050 Ala.....TyrSerAlaThrProGly.....LeuG1 2058
3879 TGATCCACTTCTAGTCTCAAGCTATGATGTGTCACAAT...GTGC 3925
      :::::
2058 yGlyGlySerArgGlyLeuThrSerPheSerHisThrGlyGlyPheThrA 2075
3926 GCACGCTGGGGATTAGCACCAACGGG...GAAGTTACTGGTGAAGAAGAG 3972
2075 rgValIleGlyAlaLeuGlyAsnThrLeuSerValValPheAspGluGlu 2091
3973 .....ACCCATTCCCTCGAAGC..... 3990
2092 PhePheArgThrLeuLeuHisArgLeuArgAlaAspValValLeuLe 2108
3991 .GTGCAA.....TACACTT 4003
2108 uValGluPheLeuLysAsnLeuProValCysLeuPheSerIleTyrThrP 2125
4004 ATGTCGAG..... 4011
2125 heileGluCysArgGlyArgIlePheProLeuArgGluPheLeuTrpGly 2141
4012 .....GAAGAGTTGCCCTCTGCAGCTGTGGCGGAAG 4046
2142 PheCysArgPheValTrpAspValIleGluGlyLeuSerIleAsnThrTy 2158
4047 ACAAGGTGATCCGTCGGGTCTGTTACCGCTGACGCTATGCTTTTCTTG 4096
2158 rGln.....TyrLeuSerAlaValAlaG 2166
4097 AAAGTGTGAAAAAAGGTGTCGAGATGCTCTTCCACCAACAGTCTAGTGGG 4146
2166 lLysValTyrArgGlyAlaPheAsnArgArgPheAsnGluGluAspGlu 2182
4147 GAAACGGCTCGTAGGTT...GAGGTGGAGCGCAAGGGTTGCTCCACAGA 4193
2183 ArgThrLeuArgValIleThrGluAlaGluGlyArgIleGluPheLeuAr 2199
4194 AACGTCGTCGTCAGCGCCGACACAAAGAAAGGGAGACGTCCAGATG 4243
2199 gValValLeuAlaGlu.....MetGluArgPheArgAlaAlaSerA 2213
4244 GTAACACAGCACAAACCGCGTCAACGAAAGCGCACAGGAGCCAGTACAG 4293
2213 snGluValSer..... 2216

```











```
3920 ATGTGGCGACGTG..... 3933
:::|::|::|::|::|
1329 spValGlnAspValLysLysGluIleLeuGluIleAlaSerSerAlaAsp 1345
3934 .....GGGATTAGCACCAACAGCGGGGAAGTTACTGGTGAGA 3968
:::|::|::|::|::|
1346 ThrProThrIleThrLysValAsnAsnLeuGlyHisLeuAspAlaLe 1362
3969 ACAGACCCATTACCTCGAAGCGTCAATACACTTATCTCGAGGAAGAGG 4018
|::|::|::|::|::|
1362 uGluAsnAspValArgHisLeuValAspIleGlnLeuArgLysGluV 1379
4019 TTGCCCGCTCTCGAGCTGTGGCGGAAGACAAAGTGATCCGTCGGGTCT 4068
|::|::|::|::|::|
1379 aIleu.....ArgMetValAspGluAspLeuAsnGluAlaLeuGluSer 1393
4069 GGTACCGGTGACGCTATGGCTTTTGTGAAAGTGTG..... 4104
|::|::|::|::|::|
1394 GlyGlyAlaGlyLysSerLysLeuCysAspLysIlePheAsnTyrIleLe 1410
4105 .....AAAAAGGTCTCG 4117
|::|::|::|::|
1410 uAspArgLeuAsnLeuPheLysHisTyrIleThrSerLysPheGlyTrpA 1427
4118 AC.....GATGCTCTTCCACCAACAGTCTAGTGGGAAACGGCTCGTGAG 4161
|::|::|::|::|::|
1427 spPheTyrAspIleLeuHis.....GluSerPheGluGlu 1438
4162 GTTAGAGGTGGCGCAAGGTTGTCTC.....CCAGAAAGCGTCTCGG 4205
|::|::|::|::|::|
1439 ValLysValArgLysSerAlaIleLeuCysAlaTyrHisAspLeu 1455
4206 TCAGGCGCGGACACAAAGAAAGGGAGAGTCACATGTTAAACACAGCAC 4255
:::|::|::|::|::|
1455 rAspIleTyrThr.....AspAsnGlyLeuValL 1465
4256 AAACCGCGGTCAACGAAGGCGACAGGACCGCAGTACAGTCCAGTCTGTG 4305
:::|::|::|::|::|
1465 euLysProIeuAsnLysSerAlaArgGluValLeuTyrAlaArgLeuIle 1481
4306 AGTTCCGCACAGGCTGATATCCAAAGGTCACCCAGTCCGAGGTACATGC 4355
:::|::|::|::|::|
1482 AlaLys..... 1483
4356 TCAGAAAGAACTGAACAAAGAGTACCATTGCGGACTGTTTCGGCGGCCA 4405
:::|::|::|::|::|
1484 .MetLysPheIleTyrArgAsnPheMetThrThrPheIle..... 1496
4406 CGCCAATCGTCGATGAGAAACCGCCCAAGTGTACG...ACTCGTGT 4452
|::|::|::|::|
1497 .....ProSerMetArgPheLeuSerGly 1504
4453 GTGAAGATAATTGACAAGGCAAGCGCTCGCTCATGTG.....GCTGA 4496
:::|::|::|::|::|
1505 ile.....LysAsnLysIleMetSerHisIleCysAspAlaG1 1517
4497 GAAAAACAGGTACAGTCGACGACCCCAACACAGAGAGTTTGACATCA 4546
|::|::|::|::|::|
1517 uLysArgGlnAlaIlePheAspLysCysLeuAspAlaThrLeuThrVal. 1533
4547 ATGAAGGCAAGCCGGTAAACAGCTTCGATGTTAGAACAGGTTCCTGCG 4596
:::|::|::|::|::|
1534 .....SerCysThrAla 1537
4597 GGTGTGCACTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTC 4646
:::|::|::|::|::|
1538 ThrAlaGlnLeu.....PheG1 1543
4647 AAACCCATTATTACCTTTCGATGAACCTTGAAGGGAGGAGTCCGCTCTTT 4696
:::|::|::|::|::|
1543 yGlyAsnPheCysLeuIleArgLeuIle.....LeuSerProIleLeuS 1558
```

```
4697 TCTCAAAGCTGGTGAGGGGTATACCTATATATGCTGAGCATGTTTCA 4746
:::|::|::|::|::|
1558 erAlaTyrLeuLysLysSerTrpThrTyrPheGlyGluHis..... 1572
4747 TCAGGCGTGCCTCGCCCTAGAGGATATCTTAACGCCAANTTAAGTACCC 4796
:::|::|::|::|::|
1573 .....AspGlyIleIleSerThrIleIleAlaSe 1582
4797 AAGCGCTCTTC.....GACCACTGTTTAGTCGAGAAGTACAAGATGGGTG 4840
|::|::|::|::|::|
1582 rThrLeuLeuGlyGlyAsnTyrSerValAsnLysMetAlaIleValS 1599
4841 GAGCGGTACCATTCACGCTGATGACGAGGAGTGC..... 4875
:::|::|::|::|::|
1599 erGlySerTyrPheHisPheLysProTyrPheCysAsnLysAlaMetLys 1615
4876 TATCCATCAGATAACCCCTATCTTGACGGTCAATCTCGTGGGGAAGCAAA 4925
|::|::|::|::|::|
1616 TyrAsnSerPheLysGlnTyrValThr...AspTyrIleAlaArgHisAs 1631
4926 CTTCTCGACTAAGTCAGGAAG..... 4947
|::|::|::|::|::|
1631 nPhePheThrThrAlaLysThrThrValGlyLeuIleThrSerProValG 1648
4947 ..... 4947
1648 lyValAlaValLeuThrValGlyAlaLeuSerGlyIleAsnProPheAla 1664
4948 .....GGTGG 4952
1665 AlaMetPhePheAlaAlaAlaSerHisTyrLysAsnPheTyrAspG1 1681
4953 TAAGTCATGGTCAATAAC.....GTAGCTTCGGGTGAC.....TATT 4990
|::|::|::|::|::|
1681 yIleValArgValSerAsnIleAlaIleAlaThrGlyAspThrLeuLeuH 1698
4991 TTCTTATGCTTTCGGTTCCTTCAAGGACGCACTTGCAATTCAGTAAACTCC 5040
|::|::|::|::|::|
1698 isLeuLysProValGlyAlaLeuLysLysAlaIleAsnLysValLysArg 1714
5041 ATCGCAGAAAGCGCATCAGTTTGACGTTCAGGCAACTCGCGCGGTCTT 5090
:::|::|::|::|::|
1715 MetLysPheLysAsnSerAsnIleThrSerAsnSerThr...ArgValAs 1730
5091 TGGTGTAGCAGGATGTTGCACGTAGCCGCGCGCTGCGGATGAGAAGT 5140
|::|::|::|::|::|
1730 nGluValAsnAsnAspGlyHisIleAlaAsnAsp...ThrAspAsnLysA 1746
5141 CACCAGGTGTTCCAAAC..... 5157
:::|::|::|::|::|
1746 snSerGluThrSerAsnIleAlaTyrGluValGluGluGlnSerAsp 1762
5157 ..... 5157
1763 AspAspGluIleValAspIleTyrMetLysArgValLysLysAspLysMe 1779
5158 .....CAGCAACCCACAGACCAAGGTGCTACCAGAA 5188
|::|::|::|::|::|
1779 tProIleLysGluTyrIleAspGlnSerGlnGluProGlyValLysGlnA 1796
5189 CAATCACACCAAAATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGG 5238
:::|::|::|::|::|
1796 laLeuThr.....AsnGlnLeuLeuLysAspSerTyrAsn 1807
5239 GAAGTCAAGGGGAGGTGCGACATACTCGATATGTCGCAACAAGATTACGT 5288
:::|::|::|::|::|
1808 AspProAsnAlaSerSerSerTyrSerThrGlyAla..... 1819
5289 TAGGAAGTGTGAGTGGCTCAGGCGCTGATAATCCAGTGTGCTCTTGAAC 5338
1820 .....V 1820
5339 CTGACTACACCCCAATG.....ACATTTGAAGTGGTT 5370
```

```
||||| ||||| .....
1820 aAlaSPHisLeuProValAspGlyLysTyrGlnSerValPheAspIleArg 1836
5371 AAACCGGGACCTCTGAAGATCGCGTGGAGTACTTGAAGTATCTGGC 5420
||||| :||| ||||| ||||| |||||
1837 LysAsnHisArgGlyProAspIleSerValGlnLeuCysGluLeuPr 1853
5421 TATA...GCATTGACAGGACATACAGGCGTTCCTATGCTAGCAATA 5467
:|:|:| ||||| ||||| ||||| |||||
1853 oMetCysGluIleMetArgAsnTyrProAla.....SerLysAsnI 1867
5468 TTGCCGCTACTACCGCGGAGGTCTCTGAAGTACTTAATCAAGTTTAT 5517
||:|:| :||| :||| ||||| |||||
1867 leValPhe.....GlnGlnThrValAsnGluValGluAsnSerIleGln 1881
5518 GAATCACTA..... 5526
1882 GluPheValGlnIleGluSerLeuThrLeuLeuLeuAsnPheLysLys 1898
5527 .....CCGGCTTTCACGTTTCAAGTCGG 5551
1898 uGluThrThrLeuAspLysPheAspSerGlyMetLysThrTyrArgGlnL 1915
5552 GCACAGATCTCATTTTTCATTCACACACAGC..... 5583
||||| ||||| |||||
1915 euGlnThrLeuValAsnAspSerThrGlnPheMetTyrValLysGlyGln 1931
5584 .....GGCTTGCCTGTGAGAGCTACCTACCGTACGTACTCAT 5618
:|:|:| ||||| ||||| |||||
1932 AspTrpValProLeuSerLeuIleHisSerAspValProGlnThrVal.. 1947
5619 AGCTGAAAGAGGTATCTTACCAAGGCGCAAGATGTCGACCGCGTGTAG 5668
||| :||| :||| :||| :|||
1948 ....GluTyrSerVal...ThrSerArgLysLeuAlaArgPheAspS 1962
5669 CTTTGGCGCAACATCTG...TTCCGTATCGCAGCATATACTGTTTCCAC 5715
:|:|:| ||||| ||||| ||||| |||||
1962 eAsnGlyAspGluValGlnPheThrThrThrGluPheAlaAlaGlyPhe 1978
5716 GATGCCAATAATTGTAGTGCTCACTGAAGTGCCTCGATCGCGCATGGT 5765
:|:|:| :||| ||||| ||||| |||||
1979 SerAsnLysLysLeuIleSerMetAspLysPheArgGlnSerAlaAsnI 1995
5766 GCGCGAATCGTTTAACTCTTCAATATAGTCTATATAGTCTCCCGCCAG 5815
:|:|:| ||||| ||||| ||||| |||||
1995 eIleAspSerAspArgLeuSerThrMetLysIleValAsnLysProPro 2012
5816 GTGGCGTAAAGACGCG.....ACGTTAGTGACGAATTCGTT 5853
||:|:| ||||| ||||| ||||| |||||
2012 lySerGlyLysThrThrGlnIleSerGlnMetMetPheAspCysHisIle 2028
5854 AAGTCACCAATAGCACGCCACCATTAACGGCTAATGTGGAAAGTCTGA 5903
||||| :||| ||||| ||||| |||||
2029 ArgGlyLysThrCysLeuAlaLeuThrSerThrLysValGlyLysAsnG 2045
5904 GGACATA.....AATATGCGGTGAAGAAGACAGATCCGAATT 5941
| ||||| ||||| ||||| |||||
2045 uLeuIleGluLysLeuLysAsnLeuIleValGlyAlaLysProLysAsnV 2062
5942 TGAAGGTCTCAACAGTGTACACAGTTAACTCCAGGTGGTAAACTTT 5991
||||| :||| :||| ||||| |||||
2062 alLysThrValAspAla.....PheAsnMetHisGlyLysLysAla 2075
5992 ATCGTCAGGGGAATGATATAAAGGTTTTGGTGATGAGTGTCACATGAT 6041
||||| :||| ||||| ||||| |||||
2076 IleValAspGluLeuTyr.....ValAspGluCysTyrMetSe 2088
6042 GCATCAAGGCTTACTACACTAGGCTCTTCGCAACCGCGCGCTCGGAAG 6091
||||| ||||| :||| :||| |||||
2088 rHisSerGlyAlaIleLeuMetAlaLeuSerAsnValAspPheThrAsnC 2105
6092 GCCTCTTTTTCGACACATATAATCAGATACCATTCATAACACGGGAGAAG 6141
||||| ||||| ||||| ||||| |||||
```

```
2105 ysLysPhePheGlyAspGluAsnGlnIleProTyrSerValArgLeuThr 2121
6142 GTGTTTAGGATGGATGTGCTGTTTTTTTCCAAAGAGAAAGCGTTGT 6191
||||| ||||| ||||| |||||
2122 GlyPheGluAlaLeuPheAlaThrThrValPheLysTyrCysHisValTh 2138
6192 ATACACTTCTTAATCTGACAGGTGTCGGTTAGATGTTTCTACTTGTGT 6241
:|:|:| ||||| ||||| ||||| |||||
2138 rHisAspSerIleSerTyrArgCysProAlaAspValCysTyrLeuLeu 2155
6242 CCTCAATGACCGTAAGGGAAACGGAAGTGTACCTGAAAGGTCGTT 6291
||| ||||| ||||| ||||| |||||
2155 eHisAlaThrAspArgAlaGlyAsnLysMetTyrThrAsnGlyValTyr 2171
6292 AGCGGTAAGGACAAACACAGTAGTAAGATCGCTGTCCAAAAGGCCAATGG 6341
:|:|:| :||| :||| ||||| |||||
2172 AlaValGlyAsnGlnAsn...IleArgThrMetHisLeuAlaLeuTh 2187
6342 AACCACTGATGACGTACCTCAATAAAGCGTGACGTGTTGTCATGA 6391
:|:|:| ||||| ||||| ||||| |||||
2187 rGlyValAlaAsnIleProIleArgGluAspThrValTyrAlaAlaPhe 2204
6392 CCAGTTGGAGAGTCCGATATGAAGAGGTCGTTGAAGGGAAGAAAGAAA 6441
||||| ||||| ||||| ||||| |||||
2204 hrGlnGlyGluLysAspGluLeuAsnArgPhe.....AlaLys 2216
6442 GAAACACCACTGATGACATGATGAAGACAGGAGGAAAAACATTACGTGA 6491
:|:|:| ||||| ||||| ||||| |||||
2217 LysAsnIleAlaLysThrValAsnGluLeuGlnGlyThrPheLysAs 2233
6492 TGTGGTATTCTTTAGGACGAGAAAGCCGATGACTCCTCTTACTAAC 6541
:|:|:| ||||| ||||| ||||| |||||
2233 nLeuIleValPheArgLeuArgLeuHisAsnValLeuTyrAsnAsp 2250
6542 AACCGCATATACTGTTGTTGTCGAGACACACACGCTCACTGGTTTAT 6591
:|:|:| ||||| ||||| ||||| |||||
2250 hrArgGlnPheIleThrAlaIleSerArgHisThrGluThrPheAsnTyr 2266
6592 GCGCTCTGAGCTCAAGTTGGACGATAGTCGGCACATATATTAGCGA 6641
:|:|:| ||||| ||||| ||||| |||||
2267 PheCysProSerAsnAlaMetAsnAspSerIleSerAsnAlaIleAsn 2283
6642 CCGCTCACCTCAATCAGTATCCGACGCTTTCACACGTTCCGCCCGG 6691
:|:|:| ||||| ||||| ||||| |||||
2283 sLeuAsn.....ThrIleGluAspAsnValValAlaGlnPheGlyPheA 2298
6692 CTGGTTGC 6699
2298 rgGlnCys 2300
seq_name: sp_virus:Q9WID7
seq_documentation_block:
ID Q9WID7 PRELIMINARY; PRT; 3115 AA.
AC Q9WID7;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE P349 PROTEIN.
OS Citrus tristeza virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=12162;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99190444; PubMed=10092023;
RA Vives M.C., Rubio L., Lopez C., Navas-Castillo J., Albiach-Marti M.R.,
RA Dawson W.O., Guerri J., Flores R., Moreno P.;
RT "The complete genome sequence of the major component of a mild citrus
RT tristeza virus isolate."
RL J. Gen. Virol. 80:811-816(1999).
RN [2]
RP SEQUENCE FROM N.A.
```



2316 CAGGTGCGCAAAAGACCCATTGCATTCATCAGGGAGTCTGGTTGTCACAGCA 2363  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1346 eLysSerCysLysSerArgValAlaLeuValIleSerGlyLysValIleHisArgA 1363  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2366 AGGTTGATATTTCTCTCATGATATGTGGGCTTTAGTTTGTCGCTGTTATG 2415  
 ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 1363 sPvalProIleAlaLeuGluTyrLeuAspGlyPheSerAlaValMetLeu 1379  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2416 GCTCAGGCCATTAAAGGATAGCGCGAAGAGTAGTATTCGCTCCCTATAACTTTAT 2465  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1380 SerAlaGlyValLysGlyArgGlnAsnAlaGluAlaPheSerArgArgLe 1396  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2466 AAAAGCCAGTAGGGGAGCTCTCGCGGGGCTTCCAAGCTCTTC...TTTC 2512  
 : | | : : : : : | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1396 uAlaAlaPheSerGlyAsp...ThrSerLeuPheGluLeuValArgPheA 1412  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2513 AGACCGTAGGCGATTGT .....TTTTCGAACCGCAGTC 2544  
 : : : : : | | | :  
 1412 lAlaIserGluLysCysArgAspLeuPheValGlyIleGlnGluAlaVal 1428  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2545 TCGGCTCATGTAAAGCAATGGTCACGATAACTTTCAACGTTTGGGAGAC 2594  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1429 GluArgCysValLysSerTyrLeuArgAsnSerPheAsnMetSerPheVa 1445  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2595 GCTTATGTCTATGCCAGAGCGTTCATCCGPAAGTACCTGGGCTCTGTTG 2644  
 : : : : : | | | :  
 1445 lAspLeuSerAspPro.....L 1451  
 : : : : : | | | :  
 2645 TTGTTACCATTGGCATTTCTCGAGCTTCAGAC...AGTTGGAGCTCAGG 2691  
 : : : : : | | | | | : : : : : | | | : : : : : | | | : : : : :  
 1451 euLeuThriIleSerGluTyrSerGluLeuAspValProIleAspLeuPro 1467  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2692 GGTGCTTTGATATTTCCAAGCAGACCTTCGGTAGGAACACTGAAGAATAG 2741  
 ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 1468 GlyPheGlyCysIleThrGluGluAspGluAlaLys ..... 1479  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2742 TCGCTTGGCGGTCTCTCTAGGGCTATCGTGAAGATTCAATTAAAG.... 2787  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1480 .....LeuIleGluAspGlyValLysSerA 1488  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2788 .....GTCATGAAGCAATGAAGACAGAA..... 2811  
 : : : : : | | | | | : : : : : | | | : : : : : | | | : : : : :  
 1488 laLeuLeuArgLysAlaValArgAspGluLaGlnLeuIleLeuGluArg 1504  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2812 .....GATGGAAAAACCCCT 2825  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1505 GlyGlyThrSerGluIleValProCysGlyAspAspaspGlyLeuGlySe 1521  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2826 GCCAATTACTGAAGATTCTGTATATGCGTTCATTAATGGGAAGCTTTCTA 2875  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1521 rAspSerSerAspSerLeuGlySer.....GlySerAspSers 1535  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2876 AGCTCCACTGTACGAGCGCAGGCTCTCTTGGCGGTTCCGAAGCAGCGGTG 2925  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1535 erSerLysAlaSerSerGlyLeuArgAlaGlyAlaAlaProSerLeu 1551  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2926 GTTTCGAGTGTTCAGGGTTTG.....GTACGTCGT.. 2958  
 |||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 1552 PheArgSerLeuLeuLysLysLeuGlyGlyPheCysValAlaArgAr 1568  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2959 .....GGCGTTCGACGAAGCCCT 2977  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1568 gLeuLeuAspLeuLeuCysGluArgLeuValGlySerLeuSerArgLeuL 1585  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 2978 TTCTCGGC.....ATTACGTGTTCTTTTCC..... 3003  
 ||| : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1585 ysPheGlyThrThrHisLeuThrSerPhePheSerMetLeuArgAlaLys 1601  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 3004 ACAGGTTTCATTATCTACGACCGCGGTTTAACTGAAGAT.....GAAG 3047  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 1602 LeuSerCysAsnPhePheAspSerIlePheThrAlaValTrpAspTh 1618  
 : : : : : | | | | | : : : : : | | | | : : : : : | | | | : : : : :  
 3048 GCTTGATGCTCTGGTGGCCACAGAAATGCTATAAACTCACCGGTGGGCA 3097

...  
1618 rValMetAsnLeuIleAArgIleValAsnGlyValThrLysThrLys... 1633  
3098 TACTGGAGACGTCGCGCGTAGCT... GTGAGCAAGGTGCTAGCT 3138  
1634 ..LeuGluCysPheArgAsnValLeuArgLeuIleProLysLeuCysSer 1649  
3139 GGAACGAA... 3147  
1650 GlyCysLysAspPheAlaLysAspPheAlaAspMetCysLysSerValLe 1666  
3148 ... GAATTTTCGACT... 3159  
1666 uLeuPheSerThrGluPhePheSerLysSerTyrThrLysLeuCysArgP 1683  
3160 ... GAAGTTTCTCTAAATGAC 3177  
1683 heSerLeuGluIlePheAspValGlyLeuIleGluIleSerAlaArgAsn 1699  
3178 TTC... ACCACTTTCCTATTGCGGAATAAGGTGCTTATCGGGAT 3218  
1700 MetGlyLeuGluThrCysValSerIleAlaSerAlaIlePheAspValII 1716  
3219 ATTCTGGCGCTCTTTGGGTGGCGGCCCAATTGCAATGCAAGTATAGGCGCG 3268  
1716 eleuThrTyProLeuGlySerIleSerPheProValIValCysAlaArgV 1733  
3269 GAATTCGGCTAAACCTAGAGGTCACGGCGACACTGATTAC... 3309  
1733 aPheGlyAlaLeuAlaMetGlu... ValGlySerAsnTyrCysPheLys 1748  
3310 ... ..GAAAC 3314  
1749 LysThrLeuGlyLysProGluIleLeuGlyLysAlaPheArgArgTh 1765  
3315 TCTAAGCTCGTTAAGTTTCACAGCGCGC... 3342  
1765 rValSerMetValLeuSerAlaLysAlaPheAspLysLeuSerValAspA 1782  
3343 ... ..GCTGTTTACCGGT 3357  
1782 laLysGlyLeuValArgCysSerGlyIleIleProGlyValLeuArgIle 1798  
3358 TTAACCTCTAGCACAGTAGTATCCGGTGGAATCTTTAGTCGTG...CG 3398  
1799 valPheSerSerLeuPheSerAspAlaAspAlaTrpValGlyTyrSerLy 1815  
3399 AAGAGGGTTTTCGTCGCGCGGTGACCGTCACT...AGGCGGA 3436  
1815 sHisGluValSerSerLeuProThrIleThrTyrSerAlaLysArgLeuS 1832  
3437 CCGTAGCTAAAGCTCAAGTCCCTTAGCGGTGCTATCGTTTCTACCTCA 3486  
1832 exPheValLeuGluGlnThrLysGluAlaCysIleAlaPheIleValAsp 1848  
3487 TACGCCATTTCCGGCTGCAGTAGTTAGGCATTTGGGCACATGCTCTTCC 3536  
1849 SerSerLysArgIleLeuAlaSerLeuThrSerGluIleHisGluSerLe 1865  
3537 ACGGCCTTAATGTGTTTCTTTTGGTTTGGGACATTCGTTGGGCGCAGGG 3586  
1865 uGlnAsnSerGluLeuThrSerGlyAlaLysSerLeuTyrGlyArgValS 1882  
3587 CTAGCGCGCAATACTTGAAGTTTCAGGCTTCTCCCAATATTTGGTGGCT 3636  
1882 erSerIleLeuSerAlaSerAsnGlyMetValAsnLysAlaGlyAsnAla 1898  
3637 GTTCCCGAGGTTGTTGGCGAGGGAAGAGTCTCAGCTCATTTGTTACTGCC 3686  
1899 ValAlaAlaValLysAlaAlaGlySerGlyLeuSerLysArgValLysAr 1915  
3687 TATTACGCTAGGGGTATCTTTTGCATCATAAAGGGGCTTGCTTTAAGCACCACCA 3736  
111

```
1915 gIleArgProSerPheHisLeu..... 1922
3737 TACCTCAACTTGTCTACGTCACCACGGTAGAGGGAGGAATGTGTACGAT 3786
1923 .....AlaArgCysValAsnGlu 1928
3787 GAGACGCTTAGTATTACGGGACATTTACTATGACGAAGGTGCTGGTCC 3836
1929 GlnThrSerGluTyrTyr..... 1934
3837 ATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGATCCA 3886
1935 .....SerA 1936
3887 CTCTTAGTGTCTCAAGCATATGATGTGTACAAATGTGCGGCACGTGGGG 3936
1936 laSerAspSerGluLeuAspGlu..SerLeuLeuSerAspThrProG1 1952
3937 ATTACCACCAACGGGAAGTTACTGGTGAAGAAGAGACCCATTACACTCG 3986
1952 yLeuArgGlyAsnGlyLys..... 1958
3987 AAGCGTGCAANTACACTTATCTCGAGGAAGAGGTTCCCGCTCTCCAGCTG 4036
1959 .....CysAsnAspLeuPhe...ArgValValSerGlyArgValLeuVal 1972
4037 TGGCGGAAGACAAGGTG.....ATCCGTCGGGTTCTGGTACCGCTGAC 4080
1973 TyrArgArgArgAlaValAlaGluAlaLeuLysArgLeuIleLeuHisValG1 1989
4081 GCTATGCGCTTTTGTGAAAGTGTGAAAAAG...GTGTCGACGATGCTTT 4127
1989 yGluCysLeuTyrSerArgIleAlaAspSerPheValArgLeuPheValA 2006
4128 TCACCAACAGTCTAGTGGGAAA..... 4150
2006 rgThrAsnThrAlaValGlyLysCysGluSerLeuAlaValAspCysAla 2022
4151 .....CGGCTCGTG.....AGGTTGAGGTGG..... 4171
2023 ThrLysIleThrAspLeuValValGlyTyrIleThrTrpArgAsnLeuH1 2039
4172 .....ACGCAAGGGTTGCTCCAGAAACGTCGTCGG 4205
2039 sPheGlyAlaTyrSerThrTrpProGlyLeuSer...GlyGlySerArgV 2055
4206 TGAGGCGCGCACAAAGAAAGGAGAGCTGCACATGGTAACACAGCAC 4255
2055 alSerLeuSerArgLeuLeuGlyAspGlyLeuThrCysValIleGlnAsn 2071
4256 AAACCGCGGTCAAGCAAGGACAGGAGCCAGTACAGTCCAGTCTTGTG 4305
2072 PheValThrAlaTrpLeuThrIlePheAsnGlu...ThrPheTyrLeuSe 2087
4306 AGTTCCGCCACAGCTGATATTTCCAAAGCTCACCCAGTCCGAGGTACATGC 4355
2087 rLeuAlaHisArgMet.....ArgThrAlaSerIleAsnValLeuL 2101
4356 TCAGANAAGAAGTGAAC..... 4372
2101 cuGluPheLeuLysAsnLeuProValCysLeuValSerIleTyrThrPhe 2117
4373 .....AAGAAGTACCATTGG..... 4387
2118 IleGluCysArgSerArgIlePhePheProLeuLysGluPheLeuTrpGlyPh 2134
4387 ..... 4387
2134 eCysIlePheValTrpAsnIleIleLysGluLeuSerValCysIleTyrG 2151
4388 .....CGACTGTTT...CGGGCGCCACGCCAA 4411
2151 luAsnLeuProValCysIleThrArgLeuCysAsnArgAlaPheArg 2167
```

---

```
4412 TCGTCGATGAGAAACCCGCCCAAGTGTTTACGACTCGTGGTGTGAAGATA 4461
2168 SerSerAsnGluAspAlaArgValThrLeuGlnValIle..... 2180
4462 ATTGACAAGGGCAAGCCGTCCTCATGTGGCTGAGAAAA..... 4501
2181 ...ThrGluAlaGluAsnArgIleGlyPheLeuArgGluValLeuAlaG 2196
4502 ....AACAGGTACAAGTCGACG.....AGCCCA 4525
2196 luMetAlaArgTyrArgAlaSerLeuGluIleProValProGlySerPro 2212
4526 AACAG..... 4530
2213 AsnSerProLeuPheProThrTyrGlyGluIleGluLeuGlyAs 2229
4531 .....AGGAGTTTGACGATCAATGAAGCAAGCGCG 4561
2229 pGlySerSerAspSerAlaArgValIleGlyValAspAsnGlySerAspS 2246
4562 GTAACAGCTTTGCATGTTTAGAACGTGTTCTCGGTCGTCAGCTGGAT 4611
2246 er.....MetSerThrGluGlyGluAlaGluLeuAsp 2256
4612 GTGTACAACGAAGCAGCTATCCGCCACAGGTTCTCAAAACGCAATTTACCTT 4661
2257 SerAspGlyAspSerAspAspSerIleProLeuAspArgAspPheGluG1 2273
4662 TGTGATCAACTTGAAGGAGGAGGTCGGTCTTTTTCACAAAGCTGGGIG 4711
2273 nIleGlyGlyLeuGlyGlySerGlyAsnValAspValPheArgLeuGlyL 2290
4712 AGGGGTATACCTATATGTTGGTAGCCAT.....GTT 4743
2290 euArgPheValTrpArgMetSerArgAsnLeuPheAsnLeuArgAlaVal 2306
4744 TCATCAGGTCGGCTCGCTCCCTAGAGGATATCTTAACGGCAATTAAGTA 4793
2307 LysPheSerTyrGlyLeuAlaLeuGlnCysIlePheAsnGlyArgSerTh 2323
4794 CCCAAGCTCTTCGACCACTGTTTAGTCAGAAAGTACAAGATGGGTGGAG 4843
2323 rGlyGlnPheIleSerArgLeuLeu.....PheSerPheGlyG:YC 2337
4844 GGTACCATTCAC..... 4857
2337 ysLeuProPheGluPheArgHisValCysAlaValArgTrpLeuSer 2353
4858 .....GCTGATGACGAG... 4869
2354 ArgSerLeuAlaLeuHisValArgAsnAlaSerIleCysAspGluGluTh 2370
4869 ..... 4869
2370 rValGluTrpLeuPheSerPheSerArgPheLeuTyrAspSerArgTyrL 2387
4869 ..... 4869
2387 euSerLeuValPheSerGluLeuGluArgValGluPheTyrLeuSerTyr 2403
4870 .....GAGTGCCTATCCATCAGATAAC..... 4890
2404 LeuPheSerLeuProLysArgCysPheThrHisAsnArgSerValGlnVa 2420
4891 .....CCTATCTTGACGTCATCTCGTGGGAAGCAAACT 4927
2420 lCysValArgAlaProValArgValValArgAsnValGlyLeuValGluA 2437
4928 TCTCGACTAAG..... 4938
2437 spGlyArgLysSerLeuSerGluValLeuAsnSerLeuAspSerLeuLys 2453
```

4939 ...TGCAGGAAGGTGCTAAGGTCATGGTCATAAACGTCAGTCTCGGTGA 4985  
 2454 valSerAsnArgLysGlyLysSerAlaValIleCysSerAspSerSeraS 2470  
 4986 CTATTTCTTTATGCTTTCGCGTGTTCCTCAAGGACGCACTTCATTCAGTAA 5035  
 2470 p.....SerGluS 2473  
 5036 ACTCCATCGAGGAGGCGGCGCATGAGTTGACGTTTCAGGCGCACTCGGCGC 5085  
 2473 erGlySerSerGlySerIleSer..... 2481  
 5086 GTCTTTGGTCTA.....GCAGGAGTCTGCAGTTAGCC...GGCGCGT 5126  
 2482 ...PheGlySerIleAsnGluGluValLeuLysThrAlaAsnGlySerVa 2497  
 5127 GTCGGATGAGAGTCACCAAGGTGTTCCAAACACGACCAACAGAGCAAG 5176  
 2497 lvalSerGluLysAla.....SerLeuMetLysHisAlaA 2509  
 5177 GTGCTACCAACAATCACACCAAAATCGGGGGCAAG..... 5214  
 2509 laAlaLeuSerGlyLeuGluProAspLeuGlyLysLysAspLeuPro 2525  
 5215 .....GCTCTATCTGAGGGAAGTGGTAGGAAGTCAA 5246  
 2526 lleIleArgArgAsnPheThrArgGlyLysSerGlyThrLysSerVa 2542  
 5247 GGGAGGTGCGACATCTCGATATGGTCCGCAACAAGATTAGTAAAGT 5296  
 2542 lGly.....SerLeuC 2546  
 5297 GTAGTGGCTCAGCGGTGATAATCCAGTATGCTGTTGAACCTGACTAC 5346  
 2546 ysgLutyrLeuThrThrLeuAsnSerSerPheGlySerValProAspLeu 2562  
 5347 ACCCAATGACATTGAAGTGGTTAAACCGGGACCTCTGAAGTCCGT 5396  
 2563 TyrProValAlaArgAspIleThr..... 2570  
 5397 CGTGGAGTACTTGAAGTATCTGCTATAGGCATTTGAGAGACATACAGGG 5446  
 2571 .....TyrLysLys.....MetThrAsnAlaMetArgC 2580  
 5447 CGTTGCTTATGGCTAGAAATATTCGCTACTACCGCGCAAGGTGTTCTG 5496  
 2580 lufPheTyrTyrSerGlnLysValThrLeuTyrGluLeuHisGlyLysLeu 2596  
 5497 AAAGTACCTAATCAAGTTTATGAATCACTACCG.....GGCTTTCAGT 5540  
 2597 .....SerSerTyrTrpAspGluLeuLysThrValGlyPheAspAr 2610  
 5541 TTACAAAGTCGGGCACAGATCTCATTTTTCATTTCAACACAGACGCGTTCG 5590  
 2610 gLysLeuAlaLysMetAspLeu.....AspAspValG 2622  
 5591 GTGTGAGACACCTACCGTAT...GTACTCATAGCTGAAAAGGATATC... 5634  
 2622 lValValAspPheAsnLeuLysValIleValGlyArgHisGlyThrArg 2638  
 5635 ...TTTACCAGGCAAGATGTCGACGCGGTGTAGCTTTGGCGCAAC 5681  
 2639 ProPheArgAsnThrValAsnSerHisGluPheMetPheCysAspG 2655  
 5682 TCTGTTTCGTATGCGAC.....GATATACTGGTTTTCACG 5716  
 2655 yLeuValLeuCysProGlyProLysLysArgAspPheAlaLeuIleSerS 2672  
 5717 ATGCCATTAAATTTATAGGTGCTACGTAAGTTCGCTCGATCGCGCATGGT 5766  
 2672 erGlnThrGlnPheValAlaAlaAsnSerPheLeuArg...AlaValAsp 2687  
 5767 GCGCAA.....TCGTTTAACTCTTCGAATATAAG.....TGCTATAA 5804

2688 GLyLysAspLeuThrPheThrAsnSerGluHisSerLeuIleValTyrGI 2704  
 5805 TGCTCCCCCAGGTGGCGTAAGACGACGAGTGTAGTGCACCAATTCGTGA 5854  
 2704 uAlaProProGlyGlyLysThrHisSerLeuValThrTyrPheAlaA 2721  
 5855 AG.....TCACCCATAGCACAGCCACCATTCAGGCTAATGTGGGAAGT 5898  
 2721 spTyrCysThrLysValSerCysLeuValThrAlaAsnLysAsnSer 2737  
 5899 TCTGAGGACATAATATGCGGTGAAGAAGAGAGATCCGAATTTGGAAGG 5948  
 2738 GluAlaGluLeuSerGlnArgIleSerArgGluPheMetAspArgLysMe 2754  
 5949 TCTC.....AACAGTGTCTACACACAGTTAACT 5974  
 2754 tLeuAlaLysHisValValLysThrAlaGlyArgValPheThrValAsps 2771  
 5975 CCAGGTGGTAACTTTATCTGTCAGGGGAATGTATAAAGGGTTTGGTG 6034  
 2771 erTyrLeuMetAsnHisLeu...ArgLeuThrThrAspLeuLeuPheVal 2786  
 6025 GATGAGGTGCACATGATGATCAAGGCTTACTACAACCTAGGCGTCTCGC 6074  
 2787 AspGluCysPheMetValHisAlaGlyAlaIleGlyAlaValValGluPh 2803  
 6075 AACCGCGCTCGAAGCGCTCTTTTGGAGACATAAATCAGATACCAT 6124  
 2803 eThrSerCysLysAlaValValPhePheGlyAspSerArgGlnIleHis 2820  
 6125 TCATAAACAGGGAGAGGTG.....TTTAGGTGGATTGTGCTGTT 6165  
 2820 yrIleHisArgAsnAspLeuGlyValSerLeuLeuHisAspIleAspAla 2836  
 6166 TTTGTTCCAAAGAGGAAACGGTTGTATACACTCTCTAAATCGTACAGTG 6215  
 2837 PheIle...GluProGlnHisArgIleTyrGlyGluValSerTyrArgCy 2852  
 6216 TCGTGTAGATTGTTGCTTACTTGTCTCAATGACCGTAAGGGGAACGG 6265  
 2852 sProTrpAspIleCysGluTrpLeuSer..... 2861  
 6266 AAAGTGTTCACCTGAAAGTGTGTAGCGGTAGGACAAACCAAGTAGTA 6315  
 2862 ..LysPheTyrProArgHisValAlaThrThrAsnThrGlySerValGly 2877  
 6316 AGA...TCGCTGTCCAAAGGCCAATTTGGAACCACTGATCAGCTAGCTGA 6362  
 2878 LysSerSerValSerIleGluValIleAsnGlyCysAspValProTy 2894  
 6363 AATAAACCGCTCAGCTGTACTTGTGTCATGACCCAGTTGGGAAGTCGGATA 6412  
 2894 rAspSerSerAlaLysTyrIleValTyrThrGlnAlaGluLysAsnGluL 2911  
 6413 TGAAGAGGTGCTTG.....AAGGAAAAGAAAAGAAACACCA 6450  
 2911 euGlnLysHisLeuGlyArgLeuThrValGlyArgThrLysAlaValPro 2927  
 6451 GTGATG...ACAGTGCATGAAGCACAGGGAACATTCAGTGTGTGTGT 6497  
 2928 ValValAsnThrValHisGluValGlnGlyGluThrTyrLysLysValar 2944  
 6498 ATTGTTTAGGACGAAGAACCCGATGATCCCTATTCCTAAACACCCG 6547  
 2944 gLeuValArgCysLysTyrGlnGluAspThrProPheCysSerAspAsnH 2961  
 6548 ATATACTTGTGGTTTTCGAGACACACAGCTCACTGGTTTATTCGCGCT 6597  
 2961 isValValValAlaLeuThrArgHisValAspSerLeuThrTyrSerVal 2977  
 6598 CTGAGCTCAAGTTGGACGATAGGTGGCGCACATATATTAGCGACGG 6645  
 6645





2102 CTCAGAGTCACCTGGA...ACTGAGGT..... 2127  
 1270 erArgValCysProArgIleSerGluValLysArgLeuArgTyrArgArg 1286  
 2128 .....ACGTATAGTCGTGGTCCCGTCGTTCTGGCGCAATC 2165  
 1287 CysAspSerAspLeuIleArgIleLysIleProArgTyrSerLysTh 1303  
 2166 CTTGGTGTTCATACCTGTTGTACCTGTTCTACTGTGTCCTTAAACAA 2215  
 1303 rArgSerCysLeuPro.....GlyCysTyrTyrL 1313  
 2216 TAGTCTCGATTCGGACTTGTTCGACAGGATCTATTCTACGCGCTCAC 2265  
 1313 euTyrLeuAspAlaLysPheValSerArgValTyrGluTyrValValAsn 1329  
 2266 ACTATAGGACATTCGAGATGAAGCTTTGAGTATGCGCTTGGCGGT 2315  
 1330 AsnCysValValValAsnAlaLysThrPheGluTrpThrTrpAsnTyrIl 1346  
 2316 CAGTGCAGAAAGACCATGTCTACAGGAGTCGCGTTGTCCACAGCA 2365  
 1346 eLysSerCysLysSerArgValIleSerGlyLysValIleHisArg 1363  
 2366 AGGTTGATATTTCTCTGATGATATGCGGTTTGTGTCGCTGTATG 2415  
 1363 spValProIleAlaLeuGluTyrLeuAspGlyPheSerAlaValMetLeu 1379  
 2416 GCTCAGCGGATTAAGGATAGGGAAGATATTCTGCTCTATAACTTTAT 2465  
 1380 SerAlaGlyValLysGlyArgGlnAsnAlaGluAlaPheSerArgArgLe 1396  
 2466 AAAGCCAGTGGAGGAGTCTCCCGGGTCTTCAAGCTCTTC...TTTC 2512  
 1396 uAlaAlaPheSerGlyAsp...ThrSerLeuPheGluLeuValArgPhe 1412  
 2513 AGACCTAGCGGATTGT.....TTTCGACGAGCAGTC 2544  
 1412 laValSerGluLysCysArgAspLeuPheValGlyIleGlnGluAlaVal 1428  
 2545 TCCGTCTATGCTAAGCAATGGTGCACGATAACTTCAAGCTTTTGGAGAC 2594  
 1429 GluArgCysValArgSerTyrLeuArgAsnSerPheAsnMetSerPheVa 1445  
 2595 GCTTATGCTATGCCAGACGCTTCATCCGTAAGTACCTGGCTCTCTTG 2644  
 1445 lAspLeuSerAspPro.....L 1451  
 2645 TTGTTACCATTTGCACCTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGT 2694  
 1451 euLeuThrIle.....SerGluTyrSerGluLeuAspVal 1462  
 2695 GCCTTTGATATTCGAAGGAGACCTTCGGTAGGAACCTGAAGAAATAGTCG 2744  
 1463 ProIleAspLeu.....ProGlyPheGlyCysIleThrGluGlyAspGI 1477  
 2745 CTTGCCGCTCTCTCTAGGCTATCTGGAAGATTCAATTAAAG..... 2787  
 1477 uAlaLys.....LeuIleGluAspGlyValLysSerAlaL 1489  
 2788 ..GTCATGAAGCAATGAACAGAA..... 2811  
 1489 euLeuArgLysAlaValArgAspGluAlaGlnLeuIleLeuGluArgGly 1505  
 2812 ...GATGAAACCCCTGCCAATTACTAGAT..... 2841  
 1506 GlyThrSerGluIleValProCysGlyAspAspGlyLeuGlySerAs 1522  
 2842 ....TCTGTATATGCTTCATATGGGGAACGTTTCTAACGTCACCTGTA 2887  
 1522 pGlySerAspSerSerLeuGlySerGlySerAspSerSerSerLysAlaS 1539

2888 CGAGGCGAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGT 2937  
 1539 erSerGlyGlyLeuArgAlaGlyAlaProSerLeuPheArgSerLeu 1555  
 2938 TCTAAGGTTTG.....GTAGTCGT..... 2958  
 1556 LeuLysLysLeuGlyGlyPheCysValAlaArgArgLeuLeuAspLe 1572  
 2959 .....GGGCTCGCACGAAGCCCTTTCTGCG..... 2985  
 1572 uLeuCysGluArgLeuValGlySerLeuSerArgLeuLysPheGlyThr 1589  
 2986 .....ATTAGCTGCTCTTTTCC.....ACAGGTTCACTA 3015  
 1589 hrHisLeuThrSerPheSerMetLeuArgAlaLysLeuSerCysAsn 1605  
 3016 TTCTAGCACCGCGTTTAACTGAAGAT.....GAAAGGCTTGATGCTCT 3059  
 1606 PhePheAspSerIlePheThrAlaValTrpAspThrValMetAsnLe 1622  
 3060 GGTCCGCACAGAGATCTATAAACTCACCGGTGGCATACTGGAGACGT 3109  
 1622 uIleArgIleValAsnGlyValThrLysThrLys.....LeuGluCysP 1637  
 3110 CGCGCGTAGCT.....GTGACGAAGTCGTAGCTGGAACGAAA... 3147  
 1637 heArgAsnValLeuArgLeuIleProLysLeuCysSerGlyCysLysAsp 1653  
 3147 ..... 3147  
 1654 PheAlaLysAspPheAlaAspMetCysLysSerValLeuLeuPheSerTh 1670  
 3148 .GAATTTTGGAGT..... 3159  
 1670 rGluPhePheSerLysSerTyrThrLysLeuCysArgPheSerLeuGluI 1687  
 3160 .....GAAGTTTCTCTAAATGACTTC..... 3180  
 1687 lePheAspValGlyLeuIleGluIleSerAlaArgAsnMetGlyLeuGlu 1703  
 3181 ACCATTTTCGTATTTCGGGAATAAGTCTTATCGGGATATTCGTGGGCTC 3230  
 1704 ThrCysValSerMetAlaSerAlaIlePheAspValIleLeuThrTyrPr 1720  
 3231 TTTGGTTCGCGCCCAATTGCATCGAAGTATAGCGCGGAATTCGCGCTA 3280  
 1720 oLeuGlySerIleSerPheProValValCysAlaArgValPheGlyAlaL 1737  
 3281 ACCCTAGAGGTACGCGGCGAGTAGTTAC..... 3309  
 1737 euAlaMetGlu...ValGlySerAsnTyrCysPheLysLysThrLeuGly 1752  
 3309 ..... 3309  
 1753 LysProGluIleLeuGlyGluAspAlaPheArgArgThrValSerMetVa 1769  
 3310 .....GAAACTCTAAGCTCGTTAAGTTCAAGCCGCGCGTGT 3349  
 1769 lLeuSerAlaLysAlaPheAspLysLeuSerValAspAlaLysGlyLeuV 1786  
 3350 TAGCGGTTTAACCTCTAGCACAGTATCCGCTGATCTTTAGTCGTGCGA 3399  
 1786 aArg.....CysSerGlyIleIleProGly.....ValLeuArg 1797  
 3400 AGAGGTTTTCGTCGCGGTGACCGTCACCTAGGCGGACCGTA..... 3441  
 1798 lIleValPheSerSerLeuPheSerAspAlaAspAlaTrpValGlyTyrSe 1814  
 3442 .....GCTAAACGT... 3450  
 1814 rLysHisGluValSerSerLeuProMetIleThrThrSerAlaLysArgL 1831  
 3451 .....CAAGTCCCTTAGCGTTGCTATCGTTTCTTACC 3483

```

1831 euSerPheValLeuGluGlnThrLysGluAlaCysIleAlaPheIleVal 1847
      |||::: ||| :::::||| ::::
3484 TCATACGCCATTTCGGCGTGCAGTATGTTAGGCATTTGGCGCATGCTCT 3533
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
1848 AspSerSerLysArgIleLeuAlaSerLeuThrSerGluIleHisGluSe 1864
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
3534 TCACGGCAGCTTAATGTTTCTTTGGTTTAGGACATGCTTGGGGCGA 3583
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
1864 rIuGlnAsnSerGluLeuThrSerGlyAlaLysSerLeuTyGlyArgV 1881
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
3584 GGGCTAGCGGATACTTGGAGATTGGAGCTTCTCCAATAATTGGTCC 3633
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
1881 alSerSerIleLeuSerAlaSerAsnGlyMetValAsnLysAlaGlyAsn 1897
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3634 GCTGTTCCGGAGCTTTGGCGAGGAAGAGTGTCCAGCTCATTTCTACT 3683
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1898 AlaValAlaIaValLysAlaIaGlySerGlyLeuSerLysArgVally 1914
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3684 GCCTATTACGCTAGGGGTATCTTTGATCATAAGGGCTTGCTTAACGACA 3733
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1914 sArgIleArgProSerPheHisLeuThr..... 1923
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3734 CCATACCTCACTTCTTACGTCCTCCACCGGTAGAGGGGAGGAATGTGTAC 3783
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1924 ..... 1935
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3784 GATGAGACGCTTAGCTATTACCGGACTTTGACTATGACGAAGGTGCTGG 3833
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1928 GluGlnThrSerGluTyTyf..... 1934
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3834 TCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATACGATGAT 3883
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1935 .....S 1935
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3884 CCACCTTAGTCTCAGCTATGATGTTGTCACAAATGTGCCGCGACGTG 3933
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1935 erAlaSerAspSerGluLeuAspGlu.SerLeuLeuSerAspThrPr 1951
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3934 GGCATTAGCACCAACGGGAAGCTTACTGGTGAAGAAGAGACCCATTCACC 3983
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1951 oGlyLeuArgGlyAsnGlyLys..... 1958
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
3984 TCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCGCTGCGAG 4033
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1959 .....CysAsnAspLeuPhe...ArgValIserGlyArgValLeu 1971
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4034 CTGTGCGGGAAGACAAGTG.....ATCCGTCGGGTCTGTGACCGCT 4077
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1972 ValTyArgArgAlaValGluAlaLeuLysArgLeuValLeuHisVa 1988
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4078 GACGCTATGGCTTTGTTCAAGTGTGAAGAAAG...GTGTCGACGATGT 4124
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1988 iGlyGluCysLeuTyfSerArgIleAlaAspSerPheValArgLeuPheV 2005
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4125 CTTTCACCAACAGCTAGTGGGAAA..... 4150
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2005 aIArgThrAsnThrAlaValGlyLysCysGluSerLeuAlaValAspCys 2021
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4151 .....CGGCTCGTG.....AGGTTGAGGTGG..... 4171
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2022 AlaThrLysIleThrAspLeuValValGlyTyfIleThrTrpArgAsnLe 2038
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4172 .....ACGCAAGGTTGCTCCAGAAAGCGTGT 4202
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2038 uIisPheGlyAlaTyfSerThrProGlyLeuSer...GlyGlySera 2054
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4203 CGGTGAGCGCGCACAGAAGAGGAGAGCTGCAGATGTTAAGACAG 4252
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2054 rGValSerLeuSerArgLeuLeuAspGlyLeuThrCysValIleGln 2070
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4253 CCAACACCGCGGTCA.....ACGAAGCGGACAGGAGCCAGTACAGTCC 4296
      :::: |||::: |||::: |||::: |||::: |||::: |||:::

```

```

2071 AsnPheValThrAlaTrpLeuThrIlePheAsnGluThrPheTyfSerSe 2087
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4297 AGTCTTGTGAGTTCCGCACAGCTGATATTCACAAAGGTCCACGACGCCA 4346
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
2087 r.....LeuAlaHisArgMet.....ArgThrAlaSerIleA 2098
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
4347 GGTACATGCTCAGAAAGAAAGTGAAC..... 4372
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
2098 snValIleuLeuGluPheLeuLysAsnLeuProValCysLeuValSerIle 2114
      :::: |||::: |||::: |||::: |||::: |||::: |||:::
4373 .....AGAAAGTACCA 4383
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2115 TyrThrPheIleGluCysArgSerArgIlePheProLeuLysGluPheLe 2131
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4384 TTGG..... 4387
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2131 utrPGLyPheCysIlePheValTrpAsnIleIleLysGluLeuSerValC 2148
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4388 .....CGACTGTGTT...CGGGCG 4402
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2148 ysIleTyfGluAsnLeuProValCysValThrArgLeuCysAsnArgAla 2164
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4403 CCACGCCAATCGTCGATGAGAAACCCGCCCAAGTCTTACGACTCGTGT 4452
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2165 PheArgArgSerSerAsnGluAspGluArgThrLeuGlnValIle... 2180
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4453 GTCAAGATAATTGACAAAGGCAAGCGCTCATGTGGCTGAGAAAAA. 4501
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2181 .....ThrGluAlaGluAsnArgIleGlyPheLeuArgGluV 2193
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4502 .....AACAGGTACAAGTCGAGC..... 4519
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2193 aIleuAlaGluMetAlaArgTyfArgAlaSerLeuGluIleProValPro 2209
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4520 .....AGCCCAACACAG..... 4530
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2210 GlySerProAsnSerProLeuPheProThrTyfGlyGluIleGluI 2226
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4531 .....AGGAGTTTGACGATCAATGAAG 4552
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2226 uLeuGlyAspGlySerSerAspSerAlaArgValIleGlyValAspAsnG 2243
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4553 GCAAGGCCGTAAACAGCTTTTCATGTTTAGAACGTTCTCCTCGGTGTG 4602
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2243 ySerAspSer.....MetSerThrGluGlyGluAla 2253
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4603 CAGCTGGATGCTACACGACGACTATCGCCACCCAGGTTCTCAACGC 4652
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2254 GluLeuAspSerAspGlyAspSerAspSerIleProLeuAspArgAs 2270
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4653 ATTACCTTTGTCGATAACTTGAAGGAGGAGTGGGCTCTTTTCTCAA 4702
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2270 pPheGluGlnIleGlyGlyLeuGlySerGlyAsnValAspValPheA 2287
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4703 AGCTGGGTGAGGGTATACCTATAATGGTGGTAGCCATGTTTTCATCAGG 4752
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2287 rGLeuGlyLeuArgPheValTrpArgMetSerArgAsnIlePheAsn... 2302
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4753 TGGCTCTGTCGCTCAGAGGATATCTTAACGGCAATTAGTACCCA...AG 4799
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2303 .....LeuArgAlaValLysPheSerTyfG 2311
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4800 CGTCTTCGACCACTGTTTAGTCAG..... 4824
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2311 yLeuAlaLeuGlnCysIlePheAsnGlyArgSerThrGlyGlnPheIleS 2328
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4825 .....AAGTACAAGATGGGTGGAGCGGTACCATTCAC... 4857
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2328 erArgLeuLeuPheSerPheGlyGlyCysLeuProPheGluPheArg 2344
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
4857 ..... 4857
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
2345 HisValCysAlaValArgTrpLeuSerArgSerLeuAlaLeuHisValAr 2361
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::

```

4858 .....GCTGATCAGCAG..... 4869  
 2361 gAsnAlaSerIleCysAspGluThrValGluTrpLeuPheAsnPhes 2378  
 4869 ..... 4869  
 2378 erArgPheLeuTyrAspSerArgTyrLeuSerLeuValPheSerGluLeu 2394  
 4870 .....GAGTG 4874  
 2395 GluArgValGluPheTyrLeuSerTyrLeuPheSerLeuProLysArgCy 2411  
 4875 CTATCCATCAGATAAC.....CCTATCTTGA 4900  
 2411 sPheThrHisAsnArgSerValGlnValCysValArgAlaProValArgV 2428  
 4901 CGGTCAATCTCTGGGGAAGCCAAACTTCTCGACTAAG..... 4938  
 2428 alValArgAspValGlyLeuValGluAspGlyArgLysSerLeuSerGlu 2444  
 4939 .....TCGAGGAAGGGTGGTAAGGT 4958  
 2445 ValLeuAsnSerLeuAspSerLeuLysValSerAsnArgLysGlyLysSe 2461  
 4959 CATGGTCATAAAGCTAGCTTCGGGTGACTATTCTTATCCCTTGGGTT 5008  
 2461 rAlaValIleGluSerAspSerAsp..... 2470  
 5009 TTCAAAGGAGCCACTTCATCTAGTAAGTAATCCATCGACGAAGGGCGCATC 5058  
 2471 .....SerGluSerGlySerSerGlySerIle 2480  
 5059 AGTTTGAGTTCAGGCAACTCGGCGCGTCTTGGTGTA.....GGCAG 5102  
 2481 Ser.....PheGlySerIleAsnGluGI 2488  
 5103 GATGTTCCAGTTAGCC...GGCGGCGTGTGCGATGAGCAAGTACCAGGTG 5149  
 2488 uValLeuLysThrAlaAsnGlySerValValSerGluLysAla..... 2502  
 5150 TTCCAAACACAGCAACACAGAGCAAGGTGCTACCAGCAACATCACACCA 5199  
 2503 .....SerLeuMetLysHisAlaAlaLeuSerGlyLeuGluPro 2516  
 5200 AAATCGGGGGCAAG.....GCTCT 5219  
 2517 AspLeuGlyGlyLysLysAspLeuProIleLeuArgArgAsnPheThrAr 2533  
 5220 ATCTGAGGGAAGTGTAGGGAAGTCAAGGGAGGTGCACATACCTCATAT 5269  
 2533 gGlyGluSerSerGlyThrLysSerValGly..... 2543  
 5270 GGTGCGCAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAAT 5319  
 2544 .....SerLeuGlyGluTyrLeuThrThrLeuAsn 2553  
 5320 CCAGTGATGCTCTGAACCTGACTACACCCCAATGACATTTGAAAGTGGT 5369  
 2554 SerSerPheGlySerValProAspLeuTyrProValAlaArgAspIleTh 2570  
 5370 TAAACCGGGACCTCTGAAGATCCGTCGTGGAGTACTTGAAGATATCTGG 5419  
 2570 r.....TyrLysLys..... 2573  
 5420 CTATAGGCATTGAGAGACATACAGGCGGTGCTTATGGCTACAAATATT 5469  
 2574 .....MetThrAsnAlaMetArgGluPheTyrTyrSerGlnLysVal 2587  
 5470 GCGGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGA 5519  
 2588 ThrLeuTyrGluLeuHisGlyLysLeu.....SerSerTyrTrpAs 2601

5520 ATCACTA.....CCGGGCTTTCACGTTTACAAAGTCGGGCACACATCTCA 5563  
 2601 pGluLeuLysThrAlaGlyPheAspArgLysLeuAlaLysMetAspLeu. 2617  
 5564 TTTTTCATTCAACACACAGAGCGGTTCGGTGTGAGAGACCTACCGTAC... 5610  
 2618 .....AspAspValGlyValValAspPheAsnLeuLys 2629  
 5611 GTTCTCATAGCTGAAAAAGTATC.....TTTACCAGGGCAAGATGT 5654  
 2630 ValIleValGlyArgHisGlyThrArgProPheArgAsnThrValAsnSe 2646  
 5655 CGACGGCGTGTAGCTTGGCGGACATCTTCCTATCGCAC..... 5697  
 2646 rHisGluPheMetPheCysAspGlyLeuValLeuCysProGlyProL 2663  
 5698 .....GATATACTGTTTCCACGATGCCATTAATTTGATAGGTGCA 5739  
 2663 yLysArgAspPheAlaLeuIleSerSerGlnThrGlnPheValAlaAla 2679  
 5740 CTGAAGTCTCGTCGATCGCATGGTGGCGAA.....TCGTTTAAAGTC 5783  
 2680 AsnSerPheLeuArg...AlaValAspGlyLysAspLeuThrPheThrAs 2695  
 5784 CTTTCAATATAAG.....TGCTATAATGCTCCCGCAGGTGGCGGTAAGA 5827  
 2695 nSerGluHisSerLeuIleValTyrGluAlaProGlyGlyGlyLys 2712  
 5828 CGACGAGCTTAGTGGACGAATTCGTTAAG.....TCACCCAATAGCACA 5871  
 2712 hrHisSerLeuValThrTyrPheAlaAspTyrCysThrLysValSerCys 2728  
 5872 GCACCATTAACGGCTAATGTGGAAAGTCTGAGGACATAAATATGCGGT 5921  
 2729 LeuValValThrAlaAsnLysAsnSerGlnAlaGluIleSerGlnArgIl 2745  
 5922 GAAGAAGAGATCCGGAAT.....TTGGAAG 5947  
 2745 eSerArgGluLeuMetAspArgLysMetLeuAlaLysHisValIleLys 2762  
 5948 GTCTCAACAGTGTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTC 5997  
 2762 hrAlaGlyArgValPheThrValAspSerTyrLeuMetAsnHisIle... 2777  
 5998 AGGGGAATGTATAAAGGTTTGGTGGATGAGTGCACATGATGATCA 6047  
 2778 ArgLeuLysThrAspLeuLeuPheValAspGlyCysPheMetValHisAl 2794  
 6048 AGGCTTACTACAACTAGCGCTCTTCGCAACCGCGCGCTCGAAGGCTCT 6097  
 2794 aGlyAlaIleGlyAlaValValGluPheThrSerCysLysAlaValValP 2811  
 6098 TTTTTCGAGACATAAATCAGATACCATTCATTAACAGGAGAGAGCTG... 6144  
 2811 hepGlyAspSerArgGlnIleHisTyrIleHisArgAsnAspLeuGly 2827  
 6145 .....TTTAGGATGGATTGCTGTTTTCCTCCAAAGAGAAAGCGT 6188  
 2828 ValSerLeuLeuHisAspIleAspAlaPheIle...GluProGlnHisAr 2843  
 6189 TGATACACTTCTAAATCGTACAGGTCTCGTGTAGATGTTTGTACTTGT 6238  
 2843 gIleTyrGlyValSerTyrArgCysProTyrAspIleCysGluTrpL 2860  
 6239 TGTGCTCAATGACCGTAAGGGAGCGAAAGTGTACCTGAAAGGCT 6288  
 2860 eSer.....LysPheTyrProArgHisVal 2868  
 6289 GTTAGGGGTAAAGACAAACAGTAGTAAGA...TCGCTGTCCAAAGGCC 6335  
 2869 AlaThrAlaAsnThrGlySerValGlyLysSerSerValSerIleGluVa 2885  
 6336 AATTGGAACCACTGATGACCTAGCTGAATAAAGCGTGCAGCTGTACTTGT 6385



1173 ysArgAlaMetValAsnLysLysAlaSerValCysTyrMetThrValVal 1189  
1858 TTCCCGCTGAGTGTGTACGGTAACGGTGAAGTTTACTTGAAGAACT 1907  
1190 ThrProGlyGluLeuLeuAspAlaArgGluSerPhePheIleLysGluLe 1206  
1908 CGATACGCTTGGTGAAGAGGAGGTGATTACCTGGCCGTACAATGTTGGTC 1957  
1206 uAspCysSerVal.....GluLeuAspThrValAlaAspArgValValT 1221  
1958 AGTGT.....GGTGAGATGATGAACATTCCTTCTCTAAGCAAGCGGG 2001  
1221 yrcysPheAsnAsnSerAlaTyrThrHisThrTyrSerThrIleCys... 1236  
2002 TTTTTCACCTTTCTTATGTACCACTTCG.....TCGGGAA 2039  
1237 .....GluTyrMetArgThrProCysValValValAspSerPh 1249  
2040 CGTGTTTAAGCTAGATGATCAGGATACCGTTCGTGGTTACCATCATCTCA 2089  
1249 eLeuPheThrIleGluMetValSerLeuArgCysSerValAsnTyrTyrC 1266  
2090 CTATGTGTAGGGCTCAGAACTACCTCGGA.....ACTGAGGTT 2127  
1266 ysIleThrLysSerArgValCysProArgIleSerGluIleLysArgLeu 1282  
2128 ACGTATAGTTCGTTG.....GTCCCGTCGTT 2153  
1283 ArgTyrArgArgCysAspSerAspLeuIleArgIleLysIleProArgTy 1299  
2154 CGTGGCAATCCGCTGCTTCATACCTGTTGTAGTGGTTCAGTGTGT 2203  
1299 rSerSerLysThrArgSerCysLeuPro.....G 1309  
2204 CCTTTAAGACAATAGCTCGATTCGGACTTTGTCGACAGGATCTATCC 2253  
1309 lYcysTyrTyrLeuTyrLeuAspAlaLysPheValSerArgValTyrGlu 1325  
2254 TAGCGCTCAACACTATAGGACATTCGAGAAATAGAACGTTTCAGTATGC 2303  
1326 TyrValValAsnAsnCysValValValAsnAlaLysThrPheGluTrpTh 1342  
2304 CGTGGGGCGGTGAGTCCGCAAGACCCATGTCATTACAGGAGTCGGC 2353  
1342 rTrpAsnTyrIleLysSerCysLysSerArgValValIleSerGlyLysI 1359  
2354 TTGTCACAGCAAGTGTATATTCCTCGATGATATGTGGGTTTAGTT 2403  
1359 leIleHisArgAspValProIleAlaLeuGluTyrLeuAspGlyPheSer 1375  
2404 GTGCTGTTATGCTCAGCGGATTAAGGATAGGCGAAGAGTATTCGCTC 2453  
1376 AlaValMetLeuSerAlaGlyValLysGlyArg...GlnAsnAlaGluAl 1391  
2454 CTATAACTTTATAAAGCCAGTCAGGGAGTCTCGCCGGGGTCTTCAAGC 2503  
1391 aPheSerArgArgLeuAla\*\*\*PheSerGlyAsp\*\*\*SerLeuPheGluL 1408  
2504 TCCTTC...TTTCAGACCGTAGGCGATTGT.....TTT 2532  
1408 euValArgPheAlaIleSerGluLysCysArgAspLeuPheLeuGlyIle 1424  
2533 TCGAAGCAGCTCTCCGCTCTATGCTAAGCAATGGTGCACGATAACTCAA 2582  
1425 GlnGluAlaValGluArgCysValLysSerTyrLeuArgAsnSerPheAs 1441  
2583 CGTTTTGGAGACCTTATGCTATGATGCCAGGCGTTTCATCCGTAAGATAC 2632  
1441 nMetSerPheValAspLeuSerAspPro..... 1450  
2633 CTGGGTCTGTTGTTAGCATTTGCACTTCTGGAGCTTCAGAC...AGG 2679  
1451 .....LeuLeuThrIleSerGluTyrSerGluLeuAspValPro 1463

2680 TTGAGAGCTCAGGGTGCCTTTGTATATTTTCGAAGAGAGACACTTCGTAGGAA 2729  
1464 IleAspLeuProGlyPheGlyCysIleThrGluGluAspGluAlaLys... 1479  
2730 ACTGAAGAATAGTCGCTTGGCGCTCTTCTCTAGGGCTATCCTCGAAGATT 2779  
1480 .....LeuIleGluAspG 1484  
2780 CAATTAAGTTCATGAAGGCAATGAAGACAGAAAGATGAAAAACCCCTGCCA 2829  
1484 lYValLysSerAlaLeuLeuArgLysAla..... 1493  
2830 ATTACTCAAGATTCTGTATATGCTGTTTCATATGCGGGAACGTTTCTAAC... 2877  
1494 ValArgAspGluAlaGlnLeuIleLeuGluArgGlyGlyThrSerGluI 1510  
2878 .GTCCACTGTACGAGGCGCAGCTCTTCTTGGC..... 2907  
1510 eValProCysGlyAspGluGluSerLeuGlyGlyAspGlySerAspSerL 1527  
2908 .....GGTTCGAAAGCGACCGTGGTTCGAGTGTTCCTAAGGGTTTGGTA 2952  
1527 euIleGlySerGlySerAspSerProSerAsnAlaSerSerGlyGlyLeu 1543  
2953 GCTCGTGGGCTCGACGAAAGCGCTTT..... 2979  
1544 ArgAlaGlyAlaAlaProSerLeuPheArgSerLeuLeuLysLysLeuG 1560  
2979 ..... 2979  
1560 yGlyGlyPheCysMetAlaLysArgLeuLeuAspLeuLeuCysGluArgL 1577  
2980 .....TCTGGCATTACGTCG 2994  
1577 euValAlaSerLeuSerArgLeuLysPheGlyThrThrHisLeuThrSer 1593  
2995 TTCTTTTTCACAGGTTCACTA.....TTCTACGAC..... 3024  
1594 LeuPheSerMetLeuArgIleLysLeuSerArgAsnPhePheAspSe 1610  
3025 .....C 3025  
1610 rIlePheThrAlaValTrpAspThrValMetAsnLeuThrArgIleValA 1627  
3026 GCGGTTTAAGTGAAGATGAA..... 3045  
1627 snGlyValThrLysThrLysLeuGluCysPheArgAsnValLeuArgLeu 1643  
3045 ..... 3045  
1644 IleProLysLeuCysSerGlyCysLysAspLeuAlaLysAspPheAlaAs 1660  
3046 .....AGCTTCATGCTCTGTCGTCACAGAGAACTCATAACTCA. 3087  
1660 pMetCysArgLeuValPheLeuLeuSerThrGluPhePheSerTyrSert 1677  
3088 .....CGGTGGGCATACG 3102  
1677 yralalysLeuCysArgPheSerLeuGluIlePheAspIleGlyLeuIle 1693  
3103 GAGACGTCG.....CGGTAGCTGTGAGC..... 3126  
1694 GluIleSerAlaArgAsnMetGlyLeuGluThrCysValSerMetAlaSe 1710  
3126 ..... 3126  
1710 rAlaIlePheAspValIleLeuThrTyrProLeuGlySerIleSerPheP 1727  
3126 ..... 3126  
1727 roIleValCysAlaArgValPheGlyAlaLeuAlaMetGluValGlySer 1743



2298 ArgAsnSerPheAsnLeuArgAlaValLysPhePheTyrGlyLeuAlaLe 2314  
4533 GAGTTTGACGATCAATGAGGCAAGCCCGT ..... 4563  
2314 uGlnCys\*\*\*PheAsnGlyGlySerMetGlyGlnPheThrSerArgLeuL 2331  
4564 ..... AAACAGCTTTGC 4575  
2331 euPheSerPheGlyAsnCysLeuProPheGluPhePheArgHisValCys 2347  
4576 ATGTTTGAACGTTCTCCGCGTGCAGCTGGATGTGTACAAC. .... 4620  
2348 SerValArgTrpLeuSerLysSerLeuAlaLeuArgValArgAsnAlaSe 2364  
4621 ..... GAAGCGACTATCCGC. .... ACCAGGTC. 4644  
2364 rIleCysAspGluGluThrValGluArgPhePheSerPheSerArgPheL 2381  
4645 ..... TCAACGCAATTTACCTTTCTCGATAACTTC 4674  
2381 euTyrAspSerArgTyrLeuSerAspValPheSerValLeuGlu\*\*Val 2397  
4675 AAAGGAGGAGTGGCTCTTTTCTCAAGCTGGTGGGGGTATACCTA 4724  
2398 AspPheTyrLeuSerTyrLeuPheSerLeuProLysArgCysLeuThrAr 2414  
4725 TAAATGGTGATACCATGTTTCATCAGGGTGGCT...CGTCCCTAGAGG 4771  
2414 gAsnArgSerValArgValCysAlaArgAlaProValArgValIleArgA 2431  
4772 ATATCTTAACGGCAATTAAAGTACCCCAAGCGTCTTCGACCACCTGTTAGTC 4821  
2431 spVal.....GlyLeuVal 2435  
4822 CAGAGTACAAGATGGTGGAGCGGTACCATTCACGCTGATCAGCAGGA 4871  
2436 GluAspGlyLysLysSer..... 2441  
4872 GTGCTATCCATCAGATAACCCCTATCTTGACGGTCAATCTCGTGGGAAGG 4921  
2442 .....LeuThrGluValLeuAsnSerLeuA 2450  
4922 CAAACTTCGACTAGTCAGCAGCAAGGTTAGTCAAGTCAATCAATCAAC 4971  
2450 spSerLeuLysValSerAspArgLysGlyLysSerAlaLeuIle..... 2464  
4972 GTAGCTTCGGGTGACTATTCTTCTATGCCTTGGCGTTTCAAGGACGCA 5021  
2464 ..... 2464  
5022 CTGTGANTCAGTAACCTCATCAGCAAA...GGGCGCATC.....AGTT 5062  
2465 ...GluSerAspAsnSerAspSerGluSerGlyArgSerPheGlySerI 2480  
5063 TGACGTTTCAGGCAACTCGCGCGCTTTGGTCTAGCAGGAGTGTGCAG 5112  
2480 leSerPheGlySerValAsnGlu.....GluValLeuGlu 2491  
5113 ...TTAGCCGCGCGTGTCCGATGAGAAGTCACCAAGGTGTTCCAAACCA 5159  
2492 AlaValAsnGlyGlyValValSerGluLysAlaPro.....Le 2504  
5160 GCAACACAGACCAAGGTCATCACCAGAACATACACCAAAATCGGGG 5209  
2504 uMetLysHisAlaAlaLeuSerArgLeuGluProAspLeuSerGlyA 2521  
5210 GCAAGGCTCTA.....TCTGAGGGAAGT 5232  
2521 snLysAspLeuProIleValArgArgSerPheThrArgGlyGluSerSer 2537  
5233 GGTAGGGAAGTCAAGGGAGGTCGACATCTCGATATGTGTGCAACAAGA 5282  
2538 GlyThrLysProAlaGly..... 2543

5283 TTACGTTAGCAAGTCTGAGTGGCTCAGGCTGATATCCAGTCTGATGCTC 5332  
2544 .....SerLeuCysValTyrLeuThrThrLeuAsnSerSerPheGlyS 2558  
5333 TTCAACTGACTACACCCCAATGACATTTTCAAGTGGTTAAACCCGGGACC 5382  
2558 erValProAspLeuTyrProValAlaArgAspIleThr..... 2570  
5383 TCTGAAGATCCCGTGGGAGTACTTTGAAGTATCTGGCTATAGGCATTGA 5432  
2571 .....TyrLysLys.....MetTh 2575  
5433 GAGGACATACAGGCGTGTCTTATGGCTAGAAATATTCGCTCCTACACCG 5482  
2575 rAsnAlaMetArgGluPheTyrTyrSerGlnLysValThrLeuTyrGluL 2592  
5483 CGAAGGTTCTCTGAAAGTACCTAACTCAAGTTTATGAATCACTACCG... 5529  
2592 euHisGlyLysLeu.....SerSerTyrTrpAspGluLeuLysThr 2605  
5530 ...GGCTTTTCACGTTTACAAAGTCGGGCACAGATCTCATTTTTCATTCAAC 5576  
2606 valGlyPheAspArgLysLeuAlaLysMetAspLeu..... 2617  
5577 ACAAGACGGCTTCGCTGTGAGAGACCTACCGTAC...GTACTCATAGCTG 5623  
2618 .AspAspAspValSerValValAspPheAsnLeuLysValIleValGlyA 2634  
5624 AAAAAAGGTATC.....TTTACCAGGCAAGATGTCGACGCGGTGTA 5667  
2634 rHisGlyThrArgProPheArgAsnThrValAsnSerHisGluPheMet 2650  
5668 GCTTTGGGCGACAATCTGTTCTGATGCGAC.....GATAT 5702  
2651 PheCysAspAspGlyLeuValLeuCysProGlyProLysLysArgAspPh 2667  
5703 ACTGGTTTCCACGATGCCATTAATTTGATAGTGCCTCAAAAGTCGCTC 5752  
2667 eAlaLeuIleSerSerGlnThrGlnPheValAlaAlaAsnSerPheLeuA 2684  
5753 GATCCGCGCATGTTGGCGAA.....TCGTTTAAGTCCTTCGAATATAAG 5796  
2684 rg...AlaValAspGlyLysAspLeuThrPheThrAsnSerGluHisSer 2699  
5797 .....TGCTATAATGCTCCCGAGTGGCGTAAGACGACGACGTAGT 5840  
2700 LeuIleValTyrGluAlaProProGlyGlyLysThrHisSerLeuVa 2716  
5841 GGACGAATTCGTTAAG.....TCACCCCAATAGCACAGCCACCATACGG 5884  
2716 LysTyrPheAlaAspTyrCysThrLysValSerCysLeuValValThrA 2733  
5885 CTAATCTGGGAAGTCTGAGGACATAAATATGCGGTGAAGAAGAGAGAT 5934  
2733 laAsnLysAsnSerGlnAlaGluIleSerGlnArgIleSerArgGluLeu 2749  
5935 CCAATTTGGAAGTCTC.....AACAGTGC 5960  
2750 MetAspArgLysMetLeuAlaLysHisValValLysThrAlaGlyArgVa 2766  
5961 TACCACAGTTAACTCCAGGCGTGAACATTTATCGTCAGGGGAATGTATA 6010  
2766 lPheThrValAspSerTyrLeuMetAsnHisIle...ArgLeuLysThrA 2782  
6011 AAAGGGTTTGTGGATGAGTGCACATGATGCATCAAGGCTTACTACAA 6060  
2782 spLeuLeuPheValAspGluCysPheMetValHisAlaGlyAlaIleGly 2798  
6061 CTAGGCGTCTTCGAAACCGCGCGTCGAAGGCTCTTTTGGAGACAT 6110  
2799 AlaValValGluPheThrSerCysLysAlaValValPhePheGlyAspSe 2815







1655 CysArgLeuValLeuProPheSer.....TThrGluPhePheSerLysSe 1669  
::: ||||| ||:::||||||| ::  
3009 T.....TACTA...TTCTACGACCGCGTTTAA 3034  
||||| ::||| |||||  
1669 rTyrThrProLeuCysArgPheSerLeuAspValPheAspValGlyLeuI 1686  
::: |||||  
3035 CTGAA.....GATGAAGCGTTGATGCTCTGTGTGCGCAGAG 3072  
||| |||||  
1686 leGluileSerAlaArgAsnMetGlyLeuGluThrCysValSerMetAla 1702  
::: |||||  
3073 AATGCTATA.....AACTCACCGETGGC 3096  
::: |||||  
1703 SerAlailePheAsnValLeuThrTyrProLeuGlySerIleSerPh 1719  
::: |||||  
3097 .....ATACTGGAGCGTCGC 3112  
::: |||||  
1719 eProValValCysAlaArgValPheGlyAlaLeuAlaMetGluValGlyS 1736  
::: |||||  
3113 CGGTAGCTGTGAGCAAGTCTGTAGCTGGAACGAAGATTTGGAGTAA 3162  
||| |||||  
1736 erAsnTyrCysPheLysSerPheGlyLysProGluIleLeuGlyGlu 1752  
::: |||||  
3163 GTTTCCTTAATGACTTCACCATTTCTGATTTGCGGAATAG..... 3204  
::: |||||  
1753 AspAlaPheArgArgThrValSerMetValLeuSerAlaLysAlaPheAs 1769  
::: |||||  
3205 .....GTGCTTA 3211  
|||  
1769 pLysLeuSerIleaspSerLysGlyLeuValArgCysSerGlyValIleP 1786  
::: |||||  
3212 TCGGATATTC..... 3222  
|||  
1786 roGlyValLeuArgIleValPheSerSerLeuPheSerAspAlaAspAla 1802  
::: |||||  
3223 .....GTGCGCTTTGGGTGGGCGGCCCAAT 3248  
|||  
1803 TrpValGlyTyrSerLysHisGluValSerSerLeuProIleThrArgVa 1819  
::: |||||  
3249 TGCATGGAATATAGCGCGGAATTCGCGCTAACGCTAGAGGTACGCG 3298  
||| |||||  
1819 lAlaArgLysSerTyrLeuSerArgTrpSerArgLeuArgLysArgAlaS 1836  
::: |||||  
3299 GCAGTAGTACGAACCT.....CTAAGCTCGTTAAGTTACAA 3336  
::: |||||  
1836 erProLeuTyrAspSerLysArgIleLeuAlaSerLeuThrSerGlu 1852  
::: |||||  
3337 .....GCCGCGGTGGTTT 3350  
|||  
1853 lIleHisGluSerPheGlnAsnLeuGluLeuThrAsnGlyAlaLysSerLe 1869  
::: |||||  
3351 ACGCGTTTAACTCTAGCACAGTATCGGTGGATCTTTAGTCGTGCGAA 3400  
||| |||||  
1869 uTyrGlyHisValSerSerValLeuSerAlaSerAsnSerValValSerL 1886  
::: |||||  
3401 GAGGTTTTTCGCGCGGTGACCGCTACTAGGCGCAC.....GTA 3441  
::: |||||  
1886 ysAla...GlyAsnAlaValAlaAlaValLysAlaAlaGlySerGlyLeu 1901  
::: |||||  
3442 GCTAAACGT..... 3450  
|||  
1902 SerLysArgValArgArgIleArgProSerPheArgLeuAlaArgCysVa 1918  
::: |||||  
3451 .....CAAGTCC 3457  
|||  
1918 lAsnGluGlnThrSerGluTyrAsnSerAlaSerAspGluSerGluLeuA 1935  
::: |||||  
3458 CTTAGCTTGTCTATCGTTTCT..... 3480  
|||  
1935 spGluSerLeuLeuSerAspThrProGlyLeuArgGlyAsnGlyLysCys 1951  
::: |||||  
3481 .....ACCTCATTACCCATTTCCGCTGCAGTATGTT 3512  
||| |||||

1952 LysAlaLeuPheArgValIleSerGlyArgValSerValCysCysArgAr 1968  
::: |||||  
3513 AGGCATTTGGGCACATGCTCTTCCACGCGACTTAATGTTTCTTTGGTT 3562  
::: ||||| |||||  
1968 gAlaVal.....GluAlaLeuLysArIleuValLeuHis.....V 1980  
::: |||||  
3563 TAGGACATTTGTTGGGCGAGGCTAGCGCAATCTACTTGGAGTTTGA 3612  
::: |||||  
1980 al.GlyAspCysLeu.....TyrThrArgIleMetAspSer..... 1991  
::: |||||  
3613 GCCTTCTCCAATAATTGTGCGCTGTTCCGAGGTTGTTGGCGAGGAA 3662  
||| |||||  
1992 .....PheValArgLeuPheValGluThrA 2000  
::: |||||  
3663 GAGTGTCA...GCTCATTTGTTACTGCCTA..... 3688  
::: |||||  
2000 snValAlaValGlyLysCysGluSerPheValValAspSerThrArgLys 2016  
::: |||||  
3689 .....TTACGCTAGGGTATCTT.....TGATCATAGG 3717  
||| |||||  
2017 lIleSerAspLeuValAsnGlyTyrIleThrTrpArgAsnLeuHisSerGl 2033  
::: |||||  
3718 GCCTTGCCTTAACGACACCATACCTCAACTTGCTTACGTCCTCCACCGGTAGA 3767  
||| |||||  
2033 yAlaTyrThrThrProGlyLeuSer..... 2042  
::: |||||  
3768 GGGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGGACCTTGACT 3817  
||| |||||  
2043 ..GlyGlySerArgSerSerLeuSerArgPheLeuAspGlyGlyLeuThr 2058  
::: |||||  
3818 ATGACGAAG..... 3826  
|||  
2059 ValCysAsnSerLysTyrCystYrArgValValAspArgPheSerMetAr 2075  
::: |||||  
3827 .....GTGCTGTGCTCCAT.....C 3839  
|||  
2075 gProLeuLeuSerLeuValHisArgMetArgThrAlaSerIleAsnValL 2092  
::: |||||  
3840 TGGGACTCAGCATGAAGCGTTCCGCTGACGATACGATGATCCACTT 3889  
||| |||||  
2092 euLeuGluPheLeuLysAsnLeuProValCysLeuValSerIleTyrThr 2108  
::: |||||  
3890 CTAGTG.....TCTCAAGCTATGATGTTGTCAAAATGTGCGCGAC 3930  
|||  
2109 LeuIleGluCysArgSerArgIlePheProLeuLysGluPhe.....Le 2123  
::: |||||  
3931 GTGGGATTTAGCACCAACGCGGAAGTTACTGTGTAAGACAGACCCATTC 3980  
||| |||||  
2123 utrpglyPheCysValPheValTrpAsnIleValLysGluLeuSerIleC 2140  
::: |||||  
3981 ACCTCGAAGCGTCAATACACTTATGTCGAGGAAGAGTTGCCCGCTCG 4030  
||| |||||  
2140 ysIleTyrGly..SerLeuProValCysValThrArg.....LeuCy 2153  
::: |||||  
4031 CAGCTGTGCGGAAACAGCAAGGTGATCGCTCGGTTCT...GGTACCGCT 4077  
||| |||||  
2153 sasn..ArgAlaPheArgArgGlySerAsnGluGlyAspValGlyThrPro 2169  
::: |||||  
4078 GACGCTATCGCTTTTGTCAAACTGTGAAAAAGGTGTCGACGATGCTTT 4127  
||| |||||  
2170 GlnValIleThrGluAlaGluAsnArgIleGluPheLeuArgGluValLe 2186  
::: |||||  
4128 TCACCAACAGCTCTAGTGGGAAACGCTCGTGTGAGGTGAGGTGCGACGCA 4177  
||| |||||  
2186 uAlaGluMetAlaArgTyrArgAlaSerLeuGluIleProValProGlyS 2203  
::: |||||  
4178 AAGGTTGCTCCGAAAGCGCTCGTGTGAGGCGCGCGACACAAAGAGG 4227  
||| |||||  
2203 er.....ProAspSerProPhe...ValProThrTyrGlyGlu 2215  
::: |||||  
4228 GGAAGAGTGCAGATGTAACACACACACAAACCGCGGTCAACGAGGCA 4277  
||| |||||  
2216 lIleGluGluLeuCysAsnAspSerSerAspSerAlaArgLeuIleGly... 2231  
||| |||||

4278 CAGGAGCCAGTACAGTCCAGTCTGTGTGAGTTCCGCCACAGGCTCATATTC 4327  
2232 |||:|||||:||||| ||| ValAspAsnGlyLeuAspSer.....ValP 2240  
4328 CAAAGGTACCCAGTCGAGTACATCTCTCAGAAAGAAAGTGAACAAGAA 4377  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2240 rothrGluGlyGluAlaGluLeuAspSerAspGlyAspSerAspSer 2256  
4378 GTACCATTTG.....GGCACTGTTTCCGGGGCCAC 4406  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2257 ileProLeuAspArgAspPheGluGlnIleGlyGlyLeuGlyGlySerG 2273  
4407 GCCAATCGTCGATGAGAAACCCGCCCAAGTGTACGACTCGTGGTCTGA 4456  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2273 yAsnVal.....AspValPheArgLeuGlyLeuL 2283  
4457 AGATAATTGACAAGGCAAGCGCTCGCTCATGTGGCTGAGAAAAACAG 4506  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2283 yspPheVal.....TtpArgMetLeuArgAsnLeuPheAsnLeuArgAla 2297  
4507 GTACAGTCGAGCAGCCCAACAGAGAGTTCACGATCAATGAAGCAA 4556  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2298 valLysPheSerTyrGlyLeuAlaLeuGlnCysIlePheAsnGlyGlySe 2314  
4557 GGCCGGT..... 4563  
2314 rMetGlyGlnPheThrSerArgLeuLeuPheSerPheGlyAsnCysLeuP 2331  
4564 .....AAACAGCTTTGTCATGTTAGAACGCTGCTTCGCGGT 4599  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2331 roPheGluLeuPheArgHisValCysSerValArgTrpLeuSerLysSer 2347  
4600 GTCGACGCTGGATGTACAAAC.....GAAGCAGCTATCGC 4634  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2348 LeuAlaIleArgValArgAsnAlaArgIleCysAspGluGluThrValG 2364  
4635 CACCAGGTTT.....TCAA 4648  
2364 uArgPheSerPheSerArgIleLeuTyrAspSerArgTyrLeuSerA 2381  
4649 ACGCATTTACCTTTGTCGATACTTGAAAGGAGGAGTGGCTTTTTC 4698  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2381 spValPheSerValLeuGluArgValAspPheTyrLeuSerTyrLeuPhe 2397  
4699 TCAAAGCTGGGTGAGGGATACCTATAATGGTGGTAGCCATGTTTCATC 4748  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2398 SerLeuProLysArgCysLeuThrArgAsnArgSerValArgValThrAr 2414  
4749 AGGTGGCCT...CGTGCCTAGAGATATCTTAACGGCAATTAAAGTACC 4795  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2414 gArgAlaProValArgValIleArgAspMet..... 2424  
4796 CAAGCGCTTCGACCACTGTTTGTAGTCACAAGTACAAGATGGGTGGAGGC 4845  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2425 .....GlyLeuValGluAspGlyLysSer..... 2433  
4846 GTACCATTCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCCTAT 4895  
2433 ..... 2433  
4896 CTTGACGGTCAATCTCGTGGGAGCAACTCTCGACTAAGTCGAGGA 4945  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2434 .LeuThrGluValLeuAsnSerLeuLeuAspSerLeuLysValSerAspArgL 2450  
4946 AGGTGCTAAGGTCAATGGTCATAAAGCTAGCTTCGGGTGCTATTTCCT 4995  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2450 ysGlyLysSerAlaLeuIle.....GluSerAsp..... 2459  
4996 ATCCCTTCGGGTTTCAAAGGAGCGCACTTGCATTCAGTAACTCCATCGA 5045  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2460 .....GlySerAspSerGluSerGlyHisSerPheGlySerIleSe 2473

5046 CGAAGGCGGCATCAGTTTGAGCTTCAGGGCAACTCGCGCGCTCTTTGGTG 5095  
2473 rPheGlySerValLys..... 2478  
5096 TAGCAGAGATGTTGCAG...TTAGCCGGCGCGGTGTCGGATGAGAAGTCA 5142  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2479 ..GluGluValLeuGluAlaValAsnSerGlyValValSerGlyLysAla 2494  
5143 CCAGGTGTTCCAAACAGCAACCAAGAGGCAAGGCTGCTACCAAGAACAA 5192  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2495 ProLeuMetLysHisAlaAspAlaLeuSerArg.....Le 2506  
5193 CACACCAAAATCGGGGGCAAG..... 5214  
2506 uGluProAspLeuSerGlyLysLysAspLeuProIleValArgArgSerP 2523  
5215 ..GCTCTATCTGAGGGAAGTGGTAGGAAGTCAAGGGGAGGTGACATAC 5262  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2523 heThrArgGlyGluSerThrGlyThrLysProAlaGly..... 2535  
5263 TCGATATGGTGGCAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGC 5312  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2536 .....SerLeuCysValTyrLeuThrTh 2543  
5313 TGATAATCCAGTGTGCTCTTGAACCTGACTACACCCCAATGACATTTG 5362  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2543 rLeuAsnSerSerPheGlySerValProAspLeuTyrProValAlaArgA 2560  
5363 AAGTGGTTAAACCGGACCTCTGAAGATGCCGTGCGTGGAGTACTTGAAG 5412  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2560 spIle..... 2561  
5413 TATCTGGCTATAGGCATTGAGAGGACATACAGGCGTGTGCTTATGGCT.. 5460  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2562 .....ThrTyrLysLysMetThrAsnAlaMe 2570  
5461 .....AGAAATATTGCGCTCAGTACCCGCCCAAGGTG 5491  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2570 tProLysPheTyrTyrSerGlnLysValThrLeuTyrGluLeuHisGlyL 2587  
5492 TTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCG.....GGCTTT 5535  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2587 ysLeu.....SerSerTyrTrpAspGluLeuLysThrValGlyPhe 2600  
5536 CAGCTTTACAAGTCGGCGCAGATCTCATTTTTCATTCAACACCAAGACGG 5585  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2601 AsnArgLysLeuAlaLysMetAspLeu.....AspAspAs 2612  
5586 CTTGCGGTGTCAGAGACCTACCGTAC...GTACTCATAGCTGAAAAAGGTA 5632  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2612 pValAspValValAspPheAsnLeuLysValIleValGlyArgHisGlyT 2629  
5633 TC.....TTTACCAAGGCAAGATGTCAGCGCGGTGGTAGCTTGGGC 5676  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2629 hrArgProPheArgAsnThrValAsnSerHisGluPheMetPheCysAsp 2645  
5677 GACAATCTGTTGTCGTATCGCAC.....GATATACTGCTTTT 5711  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2646 AspGlyLeuValLeuCysProGlyProLysLysArgAspPheAlaLeuI 2662  
5712 CCAGGATGCCATTAAATTTGATAGTGCAGCTGAAAGTCGCTCGATCGCGCA 5761  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2662 eSerSerGlnThrGlnPheValAlaAlaAsnSerPheLeuArg...Alav 2678  
5762 TGTGGGCGAA.....TCGTTTAAAGTCCTTCGAATATAAG.....TGC 5799  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2678 alAspGlyLysAspLeuThrPheThrAsnSerGluHisSerLeuIleVal 2694  
5800 TATAATGCTCCCGAGTGGCGGTGAAGACGACGACGCTTAGTGGACGAAT 5849  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: |||  
2695 TyrGluAlaArgProGlyGlyGlyLysThrHisSerLeuValThrTyrPh 2711  
5850 CGTTAAG.....TCACCCCAATAGCACAGCCACCATTTACGGCTAATGTGG 5893

```
|||||
2711 eAlaArgTyrGlyThrLysValSerCysLeuValValThrAlaAsnLysA 2728
5894 GAAGTTCTGAGCAGCATAAATATGCGGGTGAAGAAGAGAGATCCGAATTTG 5943
2728 snCysGluAlaGluLeuSerGlnArgIleSerArgGluLeuMetAspArg 2744
5944 GAAGTCTC.....AACAGTGCTACCACAGT 5969
2745 LysMetLeuAlaLysHisValLysThrAlaGlyArgValPheThrVa 2761
5970 TAACTCCAGGGTGAATACTTATCGTCAGGGAATGTATAAAGGGTTT 6019
2761 LaspSerTyrLeuMetAsnHisIle...ArgLeuLysThrAspLeuLeuP 2777
6020 TCGTCGATCAGGTGCACATGATGATCAAGCGCTTACTACAACATAGCGCT 6069
2777 heValAspValCysPheMetValHisAlaGlyAlaIleGlyAlaValVal 2793
6070 TTCGCAACGGCGCGTGGAGGCGCTCTTTTGGAGACATAAATCAGAT 6119
2794 GluPheThrSerArgLysAlaValValPhePheGlyAspSerArgGlnI 2810
6120 ACCATTCAACACGGGAGAGGTG.....TTTAGGATGGATTGTG 6160
2810 eHisTyrIleHisArgAsnAspPheGlyValSerLeuLeuHisAspIleA 2827
6161 CFTGTTTTCCTCCAAAGAGGAACCGTTGTATACACTTCTAAATCGTAC 6210
2827 spAlaPheIle...GluProGlnHisArgIleTyrGlyGluValSerTyr 2842
6211 AGGTGCGGTAGATGTTGTGCTACTTGTGTCTCTCAATGACCGTAAGGG 6260
2843 ArgCysProTTPAspIleCysGluTrpLeuSer..... 2853
6261 ACGGAAAGTGTACCTGAAGAGGTGTTAGCGGTAAAGCAAAACCCAG 6310
2854 .....LysPheTyrProArgHisValAlaThrAlaAsnThrGlySerI 2868
6311 TAGTAAGA...TCGCTGTCCAAAGGCCAATTGGAAACCACTGATCACGTA 6357
2868 leGlyArgSerSerValSerIleGluIleAsnGlyCysAspVal 2884
6358 GCTGAATAAAGCGTGCCTGCTACTGTGTCATGACCCAGTTGGAGAATC 6407
2885 ProTyrAspSerSerAlaLysTyrIleValTyrThrGlnAlaGluLysAs 2901
6408 GCATATGAAGAGTCTGTGAAGGGA...GGAAAG 6442
2901 nGluLeuGlnLysHisLeu...GlyArgLeuThrValAlaGlyThrLysA 2917
6443 AAACACCAAGTATG...ACAGTGCATGAAGCACAGCGGAAACATTCAGT 6489
2917 laValProIleValAsnThrValHisGluValGlnGlyThrTyrLys 2933
6490 GATGTGGTATTGTTTGAAGCAAGAACCGGATGACTCCCTTATTCACATA 6539
2934 LysValArgLeuValArgCysLysTyrGlnGluAspThrProSerCysSe 2950
6540 ACAACCGCATATCTGTTGTTGTCGAGACACACACGCTCAGTGGTTT 6589
2950 rAspAsnHisValValAlaLeuThrArgHisValAspSerLeuThr 2967
6590 ATGCGGCTCAGCTCAAGTTGGACGATAAGCTCGGCACATATATTAGC 6639
2967 yrSerValLeuAsnSerArgTyrAspLysThrAlaSerSerIleAsp 2983
6640 GACGGCTCACCTCAATCAGTATCCGACGCTTTGTCTCACAGCTTCGCC 6689
2984 GluAla.....ArgGluIlePheAspLysPheArgSerThrAsnHisSe 2998
6690 GGCTGGTTGC 6699
|||||
```

```
2998 rHisGlyCys 3001
seq_name: sp_virus:Q83044
seq_documentation_block:
ID Q83044 PRELIMINARY; PRT; 1873 AA.
AC Q83044;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE METHYLTRANSFERASE.
OS Lettuce infectious yellows virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=31713;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=92;
RA Klaassen V.A., Boeshore M., Koonin E.V., Tongyan T., Falk B.W.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U15440; AAA61797.1; -
DR INTERPRO; IPR000606; -
DR INTERPRO; IPR002588; -
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Methyltransf; 1.
DR TRANSFERASE; Methyltransferase.
SQ SEQUENCE 1873 AA; 217264 MW; 55A771D8357761F0 CRC64;

alignment_scores:
  Quality: 652.50      Length: 2227
  Ratio: 0.690        Gaps: 84
  Percent Similarity: 42.434 Percent Identity: 19.129

alignment_block:
US-09-301-906-3 x Q83044 ..
Align seg 1/1 to: Q83044 from: 1 to: 1873

496 GTTAAGCCCAACAGCTGTTCGGCTCCACGTACCTCTCTTAAAAACAGCA 545
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
259 ValLysProArgValSerValAsn.....LysAlaAr 270

546 AGCACTGGAGCCAGCGCAATCAGTCGCCACACAGCTGTTGGAGGAGAGG 595
: : : : : : : : : : : : : : : : : : : : : : : : : :
270 gLuvAlSerSerValGluValLysLeuGlnArgTyrGluArgGluG 287

596 CGCCTTGCAGCTTGGCCTTTCTTCAGTAAAGTGGGGGTGATGAGAGC 645
: : : : : : : : : : : : : : : : : : : : : : : : : :
287 lyGluVal..... 289

646 GACGCTGTCATCTTCGGGAAAGGAAATTTGTTTAAACAGGCGCCTTAATGT 695
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
290 ...PheValValSerLysAlaGlyLysGluAspThrArgValLeuAsnAs 305

696 TCCTATTGATGTAAGAACACAGCTTCGTTTGGCTAAAAATCTGGGATGAAG 745
: : : : : : : : : : : : : : : : : : : : : : : : : :
305 pAspAsnAlaIleArgAsnLeuPhe.....AsnA 315

746 CCTCTCGTAGGAGGGGTATTTTACGTCAAAGATAGAGCTGTTAAATTC 795
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
315 laThrLeuAsnGlyGlyLysTyr.....LysLeu 324

796 TTCCTCTATTGTCCGGGAGGCGCTACGATCGAGGACTTCATCGTGAATAC 845
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
325 HisPro..... 326

846 AGCCCCAGGGGTGTGATGTTGCCCTTCGCGCGCATTGAGTTGTGGAGTATGC 895
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
327 .....GluAlaLysSerAlaT 332

896 GCGAAAGGCGTTTGTATGCAACCACCAAGGCTGTGTGTTTAAAC... 942
: : : : : : : : : : : : : : : : : : : : : : : : : :
|||||
```

332 hrGlyLysArgPheLysTyrTyrLysAspGlyPheCysTrpLeuAspVal 348  
 943 .....AATGAGCGCTGAGGGAGAAATTACAGACGTCGTTG 980  
 349 PheAlaAspAlaAsnArgArgIle...ProGluTrpValLysProHisCy 364  
 981 CTTCCTCATCTTCCTTTTCGATAGTTCTTCGTGACGACCTTGGC..... 1023  
 364 sleuLeuThrGly.....SerValLeuMetSerCysGlyLeuTrpA 378  
 1024 ..TTTATAGCTGTTAAAGGTCATTAGGTTTCGGGGCGACGAATACAC 1071  
 378 spPheAlaLysArgLysMetValSer...ValSerHisGlyLeuLeuHis 393  
 1072 ATGCCATCACTCAATGAAGAGCCTACTTTGGGTGGAAGCGCGACAGCT 1121  
 393 ..... 393  
 1122 CTATCTCCCAATGTCCTCAAAAACCGCTATCGTCGCTGGCGATGAGACAC 1171  
 394 .TyrAspArgLysLeuGluArgSerSerAlaArgAlaGlyValArgAsp 410  
 1172 GGTGGAGGGAGAGATCTTGGCCCTCGCTCGCCCAATGCCCTTAATCAAGAG 1221  
 410 heValGly.....AlaSerAsnGluAlaValGlnArg 420  
 1222 GAGGTCTATTCATCGGTC.....GTTTCGAGTATCACCAATAGACT 1262  
 421 GluAspPheArgAspLeuArgValGluGluPheAlaGluArgVa 437  
 1263 GGT.....TTAAGGACCAATCGGCATTGCTTTCCCATTTGG 1300  
 437 lIeuGluThrAlaAsnLeuArgSerAspAsnArgLeuIleAspAsnIleL 454  
 1301 ACAGGAATTTGCGATATGTTTTCCAAAGGACGCAATGATTCGCGAA 1350  
 454 euThrArgAlaSerAspTyrIleAsnLysLysSerLysGluSerLysGlu 470  
 1351 AAACCTCATAGTGGGATGTTTTCGAAGCCGCGGAAAGGAGAA 1400  
 471 LeuAspIleAsnValCys.....LeuSerMetAspGluLysLysMe 484  
 1401 GGTGAGGGAACCTTTCCAGAGCTTTTCGATACAGTTCTCCGATCGGTCA 1450  
 484 tIleThrAsnLeuPheProAspIleGlnMetSerPheAsnGlnLysSerT 501  
 1451 GGAGTAGTCACCCATTCGCTAAATGCCATCGGAGCTGTTTCAATGGAATC 1500  
 501 yrSerAsnHisGlyValPheAsnAlaMetArgAlaCysGluAsnPheTyr 517  
 1501 TTTTCCAGGAGGTGGTAAATGTGTCTTCGATATTGGGGGAGCTT 1550  
 518 PheSerArgLysPheLysAsnSerAspTyrIleAspAlaGlyGlyAspVa 534  
 1551 CAGGTATCATGTCAAAGCTGGCCATGTGACTGTCATGTATGCAATCCAG 1600  
 534 lValSerThrLeuArgSerLysAsnHisAsnValHisIleCysSerProA 551  
 1601 TCCTAGACGTTAAAGATGTGAACGGAGATCAATGAGATCTCTTCTT 1650  
 551 rgLeuAspLeuLysAspAlaAlaArgHisIleGlnArgAlaThrValIle 567  
 1651 TCCAGAGCTGGGGAGATTGCTACGTGTCCAGTGACCTTCTTAATCAAGC 1700  
 568 AspClyLeuLysGly..... 572  
 1701 GGCTTCAAAGCTGTCTCTACTGTAGTCGAGATCGCAACTCGCAATT 1750  
 573 .TyrGlyGluThrIleSerPheCysThrAsnLysThrGluAspCysAlaV 589  
 1751 CTAGAGCCGATCGGGCTTTTATCGTGATGTGACGATATATCCCGCAG 1800  
 589 alAsnArgAspIleIleAlaValGluValTyrAspMetThrLeuArg 605

1801 CAGGTACGACAGGCTATGGATAGAAGAGGTGCGCTGTTTTCGACATAGC 1850  
 606 AspMetAlaLysAlaMetLeuSerHisGlySerArgLysPheGluPheSe 622  
 1851 TCITATGTTCCCGTGAGTGTGTGTAGCGGTAAAGGTGAAGTTTACTTGG 1900  
 622 rCysIleIleProGluIlePheThrLysGluCysAsnValGluLeuT 639  
 1901 AAGAACCTCGATACGTTGTTGAAGAGGAAGGTATTACCTGCGCTACAAT 1950  
 639 yrGluGlyArgLeuLysValThrArgIleGlyAspAsnValGluTyrTyr 655  
 1951 GTTGGTCAGTGTGTGAGATGTATGAACATTCTCTTCTTAACGTAAAGCGG 2000  
 656 TyrGlySerAsnGlyGluThrPheSerHisSerCysGlnThrLeuLysAs 672  
 2001 GTTTTTCACCTTTTATGTACGCACTTCGTCGGGAAGCGTGTAAAGC 2050  
 672 pIleLeuSerValGln...ValPheGlnPheGlyGlyArgValPheLysL 688  
 2051 TAGAGTATGAGGATACCGTGTGTGTGTACCATCATCTCATCTATGTAGG 2100  
 688 ysThrLeuGluHisSerArgGlyGlnLeuHisPhePheSerIleCysIle 704  
 2101 GCTCAGAAG...TCACCTGGAACGTGAGGTACGTATAGTGTGTTGGTCCC 2147  
 705 CysGluLysIleGluProGlySerValLysLeuLysThrTyrTyrGlnAr 721  
 2148 GTCGTTCGTGGCAAAATCGCTGTTTCATACCTGTTGTAGCTGTTCTA 2197  
 721 gSerGluLeuAspLysValThrLeuArgIleProValLysAspSerPheG 738  
 2198 GTGTGTCCTTTTAAGACAATAGTCCTC...GATTCGGACTTTGTCACAGG 2244  
 738 lYValValThrHisTyrIleIleLysGluAspArgGluPheValSerSer 754  
 2245 ATCTATTCCTACGCGCTCAACACTATAGGACATTCGAGAAATAGAAGCTT 2294  
 755 MetIleGluTyrValAlaAsnThrGlyLysIleAspLysMetVa 771  
 2295 TGATATGCGCTTGGCGGCTCAGGTGCGCAAAAGACCCATGTCATTACAG 2344  
 771 lGluTrpThrTyrSerGlnTyrArgAlaLysLysThrValThrIleLys 788  
 2345 GGAGTCGCTGTTCACAGCAAGGTTGATATTCCTCGTATGATATGTTGG 2394  
 788 erGlyLysValThrGlnLysGluThrArgIleArgLysGluLeuIlePro 804  
 2395 GGTATGTTCTGCTGTTATGTGCTCAGGCGATTAAAGGATAGGGCGAAGAG 2444  
 805 GlyPheIleAlaIleIleMetSerGluGlyIleArgAlaArgGluLysTh 821  
 2445 TATTGCTCTCTATAACTTTTATAAAGCCAGTGAGGGGAGT..... 2484  
 821 rHisTyrLeuAlaLysMetLeuTyrThrSerHisTyrLysProSerIleV 838  
 2485 .....CTCGCGGGGCTTCAAG 2502  
 838 alAsnIleIlePheArgLeuIleMetHisPheLeuGlyGlyThrLysArg 854  
 2503 CTCTTCTTTTCAGCCGTAGCGGATGTTTTCGAACGACGAGTCCCGTCTA 2552  
 855 PheIleTyrGluSerLeuValAspCysLeuLysPheLeuThr..... 868  
 2553 TGCTAAGGCAATGTCGACGATACCTCAACGTTTTCGAGACGCTTATGT 2602  
 869 .....AsnSerAspTyrIleAspThrIleValA 878  
 2603 CTATGCCACGAGCGTTTCATCCGTAAGTACCTGGGTCTGTTGTTTACC 2652  
 878 sn..... 878

2653 ATTTGCACCTTCGGAGCTTCAGACAGTTGGAGCTCAGGGGTGCCTTTGA 2702  
879 .....:|||||:|  
.....ThrGluSerArgIleGlu.....As 885  
2703 TATTTCGAAGGAGACCTTCGGTAGGAACACTGAAGATAGTCCTTCGCGG 2752  
885 pleuAspLys.....:|||||  
2753 TCCTTCTAGGCTATCGTGAAGATTCAATTAAAGTTCATGAAGCAATG 2802  
889 .....:TrpPheValPheGluGlnAsnValThrIle..... 898  
2803 AAGACAGAAGATGGAACCCCTGCCA...ATTACTGAAGATCTGTGATA 2849  
899 ...ThrThrAspAlaGluAspGlnProSerIleLeuGluGlnSerValIy 914  
2850 TCAGTTCATAATGGGGAACGTTTCTCAAGTCCACCTGTACGAGGCGAGTTC 2899  
914 sThrPheLeu.....:|||||  
2900 TTCTTGGCGGTCGAAAGCGACCGTGTTCGAGTGTTCCTAAAGGTTTG 2949  
918 .....SerLysTyrSerGluAspVal 924  
2950 GTAGTCGTGGGGCTGCGCAGC.....AAGGCTTTTCRGGCAT 2987  
925 TyrGluArgGlyAspArgSerGluPheGluLeuGluPheSer..... 939  
2988 TACGTCGTCTTTTCCACAGCTTCACTATTCTACACCGCGTTTAACTG 3037  
940 .....:|:|  
3038 AAGATCAAAGGCTTGATGCTCTGTGCGCAGACAGAAATGCTATAAATCA 3087  
941 luGlnAspLeuValAspGlnLeu.....:LeuAsnSer 951  
3088 CCGGTGGGCATACTGGAGAGCTGCGCGTAGCTGTGAGCAAGGTCGTAGC 3137  
952 Gly.....:G1 953  
3138 TCGAACGAAGAA.....TTTGGAGTCAAGTTTCTTAAATGACT 3178  
953 yGlySerLysGluAspGluTrpPheTyrSer.....:T 964  
3179 TCACCACCTTCGTATTCGCGAATAAGTGCCTATCGGGATATTCGTGCGG 3228  
964 yrIlePheSerIleValargThrLysCysThrLeuSer..... 976  
3229 TCTTTGGTGGCGGCCCAATTGCGATGGAAGTATAGGCGCGGAATT..... 3273  
977 .....LeuAlaTrpArgTrpThrAsnArgIlePheAs 987  
3274 .....GCGGCTAACGCTAGAAGTACCGGCGCAGTAGTT 3307  
987 nPheIleThrSerCysArgAlaLysGlyLysAspPheValLysTyrIleT 1004  
3308 ACGAAACTTAAGCTCGTAAAGTTCACAAAGCGCGCGGTTCAGCCGT 3357  
1004 yrGluIleIleLysIleIleIleAsnLysPro.....:TrpGluAla 1017  
3358 TTAACCTCTAGCACAGTATCCGGTGGATCTTTTAGTCGTGCGAAGAGGTT 3407  
1018 LeuCysHisTyrLeuValGlnGlyTyrLysThrValLeuLys..... 1031  
3408 TTCGTGCGCGGTGACCGTCACTAGGCGCACCGTACGTAACGTCAGTCC 3457  
1032 .IleProMetThrCysValAspSerValAsnTyrLeuLysArgLysSerL 1048  
3458 CCTTAGCTTGCTATCGTTTTC.....:TAC 3482  
1048 euGluLysIleLeuSerTyrPheSerHisGluArgAlaAspAspLutyr 1064  
3483 CTCATACGCCATTTCCGCGCTGCAGTATGTTAGGCATTTGGGC..... 3524

1065 IlePheProGlyTyrAspSerAspPheSerGluGluSerGlyAspTyrG1 1081  
3525 .....:|:|  
.....ACATGCTCTTCCACGGCAC 3543  
1081 yAspGluAsnThrGluLeuAspSerArgSerArgCysThrValIleLysS 1098  
3544 TTAATGCTTTTCTTTGGTTTAGGACATTGCTTGGGGCGAGGGCTAGCG 3593  
1098 erLysValGluMetLeuPheArgAsnIle.....:Lys 1108  
3594 GAATACCTTGGAAAGTTTGGAGGCTTCTCCAATAATTGGTGGCGTGTTC... 3641  
1109 GlyHisIleGluLysPhe.....:LeuGlnArgCysGlyI1 1120  
3642 .....:CGAGTGTGTTGGCGA 3657  
1120 eIleAspGlnTyrAsnLysIleTyrHisTrpPheLysSerIleLeuThra 1137  
3658 GGAAGAGGTGTCAGCTCATTTGTTACTGCTTATACGCTAGGGGTATCTTT 3707  
1137 spValAspCysPheAsnSerValLeuAsnPhe.....:Ile 1148  
3708 GATCATAAGGGCTTGTTCACGACACCATACCTCAACTTGTCTTACGTCC 3757  
1149 ThrHisThrGly..Cys.....AspAlaLeuLeuAlaValLeuThrg.y 1162  
3758 CACCGGTAGAGGGGAGGAATGTGTACGATGACAGCTTAGGTATTACCGG 3807  
1163 HisPhe...SerIleAlaSerCysAlaLeuLys..... 1172  
3808 GACTTTGACTATGACGAAGGTGCTGCTCATCTGGGACTCAGCATGAAGC 3857  
1173 .....PheMetThrAspValIleIleGluLysAsnLeuSerGluLysH 1187  
3858 GGTTCGCGGTGACGATACGATGATTCCTTCTAGTGTCTCAAGCTATG 3907  
1187 sHisThrThrLysLeuIleAlaGlyAsnLeuThrGlnCysIleTyr 1203  
3908 ATGTGTGTCA.....CAAATGTCCGCGACGTGGGATTAGC 3942  
1204 LysLeuAspPheIleHisProIleTrpIleProIleArgTrpGlyVally 1220  
3943 ACCAAGCGGGAAGTTACTGGTGAAGAAGACCCATTACCTCGAAGCGT 3992  
1220 sGluIleSerLysIleHisIleLysLysLysLeuLeuAsn..... 1233  
3993 GCAATACACTTATGTCGAGGAAGAGTTGCCCGCTGTGCAGCTGTGGCGG 4042  
1234 .....:TrpArg 1235  
4043 AAAGACAAGGTGATCCGTGGGTTCGTGCTACCGCTACCGCTATGCTTTT 4092  
1235 ..... 1235  
4093 GTTGAAGTGTGAAAAAGGTGTCGACGATGCTTTTACCAACAGCTAG 4142  
1236 .....SerThrLysGluValThrAsnGluSerI 1245  
4143 TGGGGAACCGGCTCGTGAGGTTGAGGTGACGGCAAAAGGTTGCTCCAG 4192  
1245 LeCysLysAspLeuVal..... 1250  
4193 AAAGCGTCTCGTGGTGGCGCGCACACAAGAAGGGAAGAGCTGCAGAT 4242  
1250 ..... 1250  
4243 GGTAAACAGCACAAACCGCGGTCAACGAAGCGACAGGGAGCCAGTACA 4292  
1251 .....HisLys.....SerTyrIleLysTyrP 1258  
4293 GTCCAGTCTTGTGAGTTGCGCCACAGGCTGATATTCAAAGGTCAACCCAGT 4342  
:|:

1258 heAspIle..... 1260  
4343 CCAGGTACATGCTCAGAAAGAACTGAACAAGAAGTACATTGGCGACT 4392  
1261 .....LysLysAlaValTrp..... 1265  
4393 GTTTGGGGCCACGCCAATCGTGCATGAGAAACCCGCCCAAGTGTAC 4442  
1265 ..... 1265  
4443 GACTCGTGTGTGAAGATAATTGACAAGGGCAAGCGCGTCGCTCATGTGG 4492  
1266 .....LeuLeuTrp 1268  
4493 CTGAGAAAAAACAGGTACAAAGTCGAGAGCCCAACAGAGGAGTTTGACG 4542  
1268 ..... 1268  
4543 ATCAATGAAGCAAGCGGCTTAACAGCTTTGCATGTTTAGA..... 4584  
1269 .....PheSerLeuLeuIleIlePheL 1276  
4585 .....ACGTGTCTCGCGTGTGCGAGCTGGATGTGTACAAACGAAGCG 4626  
1276 euHisProThrLeuGlyPhePheIleLeuPheSerIleIleThrProAlaIle 1292  
4627 ACTATGCCCAACAGGTCTCAACAGCATTTACCTTTGTGCATAACTTGAA 4676  
1293 GluIleLysArgTyrTyrAsnAsnValValCysPhe...AsnAsnIleIle 1308  
4677 AGGAGGAGTCGCGCTTTTCTCAAGCTGGGTGAGGGGTATACCTAT. 4725  
1308 sMetSerTyrProHisValIleAspArgLeuGlyThrTyrAsnPheA 1325  
4726 .....AATGGT 4731  
1325 laLysLeuLysLysAlaValArgGluLysPheSerGluAsnLysSerGln 1341  
4732 GTTAGCATCTTTTCATCAGGGTGGCTCGTGGCCCTAGAGATATCTTAAC 4781  
1342 GlyValIleSerGluIleProAlaGluLysAspGluAspValMetPr 1358  
4782 GGCAATTAAGTACCAAGCGCTCTCGACCACTGTTTAGTCAGAGTACA 4831  
1358 oProLeuGluGlu.....ValGluThrProL 1367  
4832 AGATGGGTGAGCGGTACCATTCACCGCTGATCAC.....GAGGAG 4872  
1367 ysArgSerValGlyArgProSerSerLysGlnAspIleGluAspGlu 1383  
4873 TGTATCCATCAGATAACCTATCTTTGACGGTCAATCTC..... 4911  
1384 CysAsnIleLysGluAspProThrGluSerLeuAsnPheSerAsnIleG1 1400  
4912 .....GTGGGAAGGCAAACTTC. 4929  
1400 uAlaThrSerAsnArgTyrAspThrLeuArgValGlyAsnSerLysLeuC 1417  
4930 .....TCGACTAAGTCGAGGAAGGTGCTAAG 4956  
1417 ysLysPheLeuArgPheTyrProLysSerLysCys..... 1428  
4957 GTCATGGTCATAAACCTAGCTTCGGGTGACTATTTTCTTATGCTTCCCG 5006  
1429 .....TyrAlaAsnIleGlnThrGlyAspLeuIle..... 1438  
5007 TTTTCAAGGACCCACTTGCATTTCAGTAACCTCCATCGACGAGGGCGCA 5056  
1439 .....ArgAspSerIleGluGluPheTyrHisLeuGluArgSerLysL 1453  
5057 TCAGTTTACGTTCAGGGCAACTCGCGCGCTCTTGGTGTA..... 5097  
1453 euAspIleGluIleSerLysMetGlnArgValValGluIleMetAsnAsp 1469

5098 ...GGCAGGATGTTTCAGTTAGCCGGCGCGGTGTCGGATGAGAGTCACC 5144  
1470 ThrGlyArgMetLeuAspAsnValArgArgMetIleAspArgSerMe 1486  
5145 AGGTGTTCCAAACAGCAACACACAGAGCCAAAGGTGCTACCAGAACATCA 5194  
1486 tTyrVal..... 1488  
5195 CACCAAAATCGGGGGCAAGGCTCTATCTGAGGAAAGTGTAGGGAAGTC 5244  
1489 .....ThrLeu 1490  
5245 AAGGGAGGTGACATACTCGATATGTGCGAACAAAGATTACGTTAGGAA 5294  
1491 AspGlyValSerTrpTyrArgLeuGlyCysLysAsp..... 1502  
5295 GTGTGAGTGGCTCAGGCGCTGATATCCAGTATGGCTCTTGAACCTGACT 5344  
1502 ..... 1502  
5345 ACACCCCAATGACATTTGAAGTGGTTAAACCGGGACCTCTGAAGATGCC 5394  
1503 .....LysAsnProAlaLysGluAsp... 1509  
5395 GTCGTGGAGTACTTGAAGTATCTGGCTATATAGGCATTGAGAGGACATACAG 5444  
1509 ..... 1509  
5445 GCGTTCCTTATGGCTAGAAATATTGCCGTCACCTACCCCGGAGGTGTC 5494  
1509 ..... 1509  
5495 TGAAGTACCTTAATCAAGTTTATGAATCACTACCGGGCTTTACGTTTAC 5544  
1510 .....Phe 1510  
5545 AGTCGGGCACAGATCTCATTTTTCATTCAACACAGACGGCTGCGGT 5594  
1511 LysAlaIlePheAspHisAsnPheAsnIleMetGluGlyAsnAlaLysVa 1527  
5595 GAGAGACCTACCGTACGTACTCATAGCTGAAAAGGTATCTTTACCAAGG 5644  
1527 lLysGluPheAlaValSerSerAspGluTrpArgGly..MetTyrSerAs 1543  
5645 GCAAAGATGTCGACGGGTGGTAGCTTTGGCGCACAACTCTGTCGTATGC 5694  
1543 nGluArgCysArg..AlaIle.....GluGlnLeuPheAsnAsp 1555  
5695 GACGATATACTGTTTTCACGATGCCATTAAATTTGATAGTGCACGTAA 5744  
1556 AsnAsnGluMetValArgArgProAspValAsn..... 1566  
5745 AGTCGCTCGATCGGCATGGTGGCGAATCGTTTAAGTCTCTCGAATATA 5794  
1567 .....GlyLeu.....L 1569  
5795 AGTCGTATAATGCTCCCGCAGGTGGCGTAAGCAGCAGCGTGTAGTGCAC 5844  
1569 ysPheTyrAsnLysProGlyAlaGlyLysThrThrIleAlaLys 1585  
5845 GAATTCGTTAAGTCAACCAATAGCACAGCCCAATTCAGGCT..... 5886  
1586 LeuMetSerLysAspLeuLysAsnLysValLysCysLeuAlaLeuSerTy 1602  
5887 .....AATGTGGGAAGTCTCGAGCACATAAATATG.....CGGG 5920  
1602 rThrLysValGlyArgLeuGluLeuIleAspLysLeuLysLysaspGlyI 1619  
5921 TGAAGAAGACAGATCCGAATTTTGAAGGTCTCAACAGTGTCTACCACAGTT 5970  
1619 leGluLysProGluLysTyrValLysThrTyrAspSerPheLeuMetAsn 1635

```
5971 AACTCCAGGTCGTAACACTTTATCGTCAGGGGAATGTATATAAAGGTTT 6020
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1636 AsnAspAsnIleLeuGluIleVal.....AsnLeuTy 1646

6021 GGTGATGAGGTGCACATGATCATCAGGCTTACTACACTAGGCGTCT 6070
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1646 rCysAspGluValPheMetMetHisAlaGlyHisPheLeuThrLeuLeuT 1663

6071 TCGCAACCGCGCGTCGGAAGGCCCTCTTTTGGAGACATAAATCAGATA 6120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1663 hrLysIleAlaTyrlGlnAsnGlyTyrcysTyrglyAspValAsnGlnIle 1679

6121 CCATTCTATAACAGCAGCAGAGGTGTAGGATGGATGTCTGCTGTTTGT 6170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1680 ProPheIleAsnArgAspPro.....TyrTh 1688

6171 TCCTCA.....AGAAGGAAGCGTGTGTATACA 6196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1688 rProAlaTyrlLeuSerArgGluPhePheArgLysGlnAspLeuAsnTyra 1705

6197 CTTCTAAATCGTACAGGTCTCGGTAGATGTTTGTCTACTTGTGTCTCA 6246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1705 spThrTyrlTyrlArgCysProLeuAspThrCysTyrlLeuLeuSerAsn 1721

6247 ATGACC.....GTAAGGGGAAC 6263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1722 LeuLysAspGluMetGlyAsnIleIleTyrlAlaGlyGlyValLysAsnVa 1738

6264 GGAAGAGTGTACCTGTAAGAGTCGTTAGCGGTAAAGCAACACAGTAG 6313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1738 lAsnGluValTyrl.....ProThrI 1745

6314 TAAGATCGCTGCCAAAGGCCAATTGGAACCACTGATGACGTA...GCT 6360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1745 leArgSerLeuAsnLeuPheGlyIleAsnValValGlyLeuProVal 1761

6361 GAAATAAACCGCTGACGTGTCTACTGTGCATGACCCAGTTGGAGAAGTCGA 6410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1762 GluTyrlAsnAlaLys...TyrlLeuThrPheThrGlnAspGluLysLeuAs 1777

6411 TATGAAGAGTGTGTGAGGGGAAGGANA...GAACACACAGTGTATGA 6457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1777 nLeuGlnArgHisIleAspSerGlnGlyCysArgAsnAlaValSerT 1794

6458 CAGTCGATGAACACAGGAAAAACATTCAGTGATGTGCTATTGTTAGG 6507
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1794 hrValAsnGluAlaGlnGlyCysThrPheSerGluValAsnLeuValArg 1810

6508 ACGAAGAACCGCATCTCCCTATTTCATCAACACCGCATATACTTGT 6557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1811 LeuValGlnPheAspAsnProValMetSerAspIleAsnGlnPheValVa 1827

6558 TGGTTTGTGACACACACACGCTCACTGGTTTATCGCGCTGTGAGCTCAA 6607
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1827 lAlaIleSerArgHisThrThrPheLysTyrlPheThrProHisSerA 1844

6608 AGTTGACGATGAAGTCGSCACATATATTAGGACGCGCTCACCTCAATCA 6657
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1844 rLeuAsnAspArgValSerAsnAlaIleSer.....SerLeuGlnSer 1858

6658 GPATCCGAGCTTTGCTTCACACGTTTC 6684
|||||:|||||:|||||:|||||:|||||:|||||:
1859 ValSerAspPheValLeuLysAspTyrl 1867
```

seq\_name: sp\_virus:Q9JGK3

```
seq_documentation_block:
ID   Q9JGK3
AC   Q9JGK3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   152-KDA PROTEIN.
```

```
OS   soil-borne wheat mosaic virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
OX   NCBI_TaxID=28375;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   STRAIN=JAPANESE;
RT   "Similarity and divergence among viruses in the genus Furovirus.";
RL   Virology 270:201-207(2000).
DR   EMBL; AB033689; BAA94796.1; -.
SQ   SEQUENCE 1334 AA; 151950 MW; 288761E57551D0A7 CRC64;
```

```
alignment_scores:
  Quality: 486.00      Length: 1919
  Ratio: 0.606        Gaps: 82
Percent Similarity: 41.793 Percent Identity: 19.854
alignment_block:
US-09-301-906-3 x Q9JGK3 ..
Align seg 1/1 to: Q9JGK3 from: 1 to: 1334
1192 GCCTCCGTCGCCAATCCCTTAATCAAGAGGAGGTCTATTTCATCGGTCT 1241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 SerAsnIleLeuGlyIleIleAsnGluGluValValArgAlaA1 22
1242 TTCGAGTATCACCAATAGACTGGTATTAGGACCACATCGGCATTGCTTT 1291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 ealThrSerAlaThrArg.....ThrAsnS 31
1292 CCATTTTGGACACAGAAATTTGCGATATGTTTCTCAAAGG .....GAC 1335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 erGluLeuHisArgThrValCysGluGlnIleArgGluGlnPheLeuAsp 47
1336 GCAATGATTTCGCGAAAAACCCCTCACATATGCGATGTG.....TTTCT 1379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 ThrValGluAsnGlnLysGlnLysLysIleAspValArgArgAspLe 64
1380 GAAGCGCGGGAAGGAGAGACTGAGGAACTCTTCCAGAGCTTTCGA 1429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 uThrGlnGluGlnLeuGlnThrLeuAsnGluLeuTyrlProGluArgHisI 81
1430 TACAGTTCTCCGACTCGGTGAGGTAGTCACCATTCGCTAATGCGCATG 1479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 leValThrSerGlyCysGluArgGlyThrHisSerPheAlaAlaLeSer 97
1480 CGAGCTGTTTCAATGGA...ATCTTTTCCAGGAGGTGCGTAATGTGTG 1526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 ArgLysIleGluThrAspLeuLeuLeuSerArgMetProLysAsnAlaAs 114
1527 CTTCTTCGATATTGGGGAGGCTTCACGTATCATGTCAAGCTGGCCATG 1576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 nValTyrlAspIleGlyGlyAsnTrpAlaThrHisLeuLysArgLysAspT 131
1577 TG...AAGTGTCTATGATCAATCCAGTCTAGACGTTAAAGATGTGAAG 1623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 hrArgLysValHisCysCysProValLeuAspPheArgAspAlaGln 147
1624 CGGAGAATCAATCAGATCCTC.....TTTCTTTCCACAGCTGG 1661
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 ArgLysThrThrArgTrpLeuSerValGluLysPheLeuSerGluLysAs 164
1662 GGA.....G 1666
|||||:|||||:|||||:|||||:|||||:|||||:
164 pSerIleSerGluHisCysGlyGlnLysValLeuGluIleGlnGluAspG 181
1667 ATTGCTACGTGTCCAGTGACCTTCTAACTGAAGCG...GCTTCAAAGTCT 1713
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 luAspArgIleSerAsnAsnLeuArgLysGlyAlaValAlaProGluAsp 197
1714 GTGTCT.....TACTGTAGTCAGAAATCGCAGAACTCGCATTTCTTAG 1754
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```





3461 TAGCGTTGCTATCGTTTCTTACCTCATACGCCAATTCGCGGTGCAGTATG 3510  
603 .....  
3511 TTAGGCATTTGGGCACATGCTCTTCCAGGCCACTTAATGTTTCTTTGG 3560  
603 .....  
3561 TTTAGGGACATTGCTTTGGGCGAGGCTAGCGGAATACTTGAAGTTTG 3610  
604 .....PheT 605  
3611 GAGGCTTCCCAATAATATGGTGGCGTGTCCCGAG.....GTT 3648  
605 hrAsnPheValAspGluTrpCysAsnLysGluAspHisPheAsnHisVal 621  
3649 GTTTGGCGAGGGAAGAGTGTCTAGCTCATCTGTTACTGCCTATATACCTAGG 3698  
622 HisIleGluValGluAsnLysTyrSerLeuLeuValLysIleValIleAl 638  
3699 GGTATCTTTGATCATAGGGGGCTTCTTAACGACACCATACCTCAACTTG 3748  
638 aVal.....TrpGluA 642  
3749 CTTACGTCCTCCACCGGTAGAGGGAGGAATGTGTACGATCAGACGCTTAGG 3798  
642 IaIleValProMetGlnPheAlaProValTyrIleAsp..... 655  
3799 TATTACCGGACTTTGACTATGACCAAGGTGCTGTCCATCTGGGACTCA 3848  
656 .....AspGluGluArgAspProSerAspAlaG1 665  
3849 GCAT.....GAAGCGGTTCCGGTGACGATAACGATGATCCACTTCTA 3892  
665 uSerValValGluLeuAspLeuSerGluThrAspAspGlySerGlyLysG 682  
3893 GRTCTCAAGCTATGATGTTGTACAAATGTGCGCGACGTGGGGATTAGC 3942  
682 Ly.....AspTyrAlaGluLys 687  
3943 ACCACGGGGAAATTACTGGTGAAGAAGACGACCCATTACCTCGAAGCGT 3992  
688 ProGlySerValGluLysAlaGluLysProIleGluAlaArgValSerte 704  
3993 GCAATACACTTATGTCGAGGAAGAGTGTGCCCGCTCGACGTGTGGCGG 4042  
704 uGluLysGlyLysGluLysValGluValGluAspSerAlaIle..... 718  
4043 AAAGACAAGGTGATCCGTCGGGTCTGTGTACCGTGCAGCTATGCTTTT 4092  
719 ..ArgArgGly.....SerGlySerSerSerProAspSer..... 729  
4093 GTTGAAGTGTCAAAAAGGTTGTCACGATGCTTTCCACACAGCTCTAG 4142  
729 ..... 729  
4143 TGGGGAACGGCTCTGAGTTGAGGTG.....GACGGCAAAAGGTTGC 4186  
730 .....GluThrSerMetGluAspGluLeuValSerAspGlyGluSerLeup 745  
4187 TCCCAAGAAGCGTCTCGGTGAG.....GCGCCGACACAGAAGGGGA 4230  
745 heGlyArgSerCysGlnAlaThrIleTyrSerSerThrSerLeuTyrGly 761  
4231 AGAGTGCAGATGTTAACACAGCACAAACCGGTCACAGAGGCGACAG 4280  
762 SerSerGluAspAspThrSerValGlySerAlaSerSerGlu..... 775  
4281 GGAGCCAGTACAGTCCAGTCTGTGTGAGTTCGCCACAGCGCTGATTTCCA 4330  
776 .....IleAspSerArgValLeuGluAlaValGluAlaIleAspG 790  
4331 AGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAGTCAAAAGCTA 4380

790 LuMet.....GluLysGluValSerAlaGluVal 799  
4381 CCATTGGCGACTGTTTCGGCGCCCGCCCAATCTCGTCGATGAGAAACCCGC 4430  
800 LeuAlaIaIaValVal.....ValValThrGluSerSerG1 811  
4431 CCCAAGTGTACGACTCGTGTGTGAAGATAATTCACAAGGGCAAGGCCG 4480  
811 uProGluValSerGluValValVal.....GluAsnA 822  
4481 TCGCTCATGTGCTCGAGAAAAACAGGTACAAAGTCGAGCAGCCCAACAG 4530  
822 IaAlaThrValGluCysLysGluGlnValValValAspThrThrAspSer 838  
4531 AGGAGTTTGACGATCAATGAAGCAAGGCCGTAAACAGCTTTGCATGTT 4580  
839 GluSer.....AspGluGluArgAlaLys..... 846  
4581 TAGAACGTGTTCTCGGTGTGTCAGCTGATGTGTACAAACGAAGCGACTA 4630  
847 .....SerTrpGly.....SerMetAlaGluGluGluSerA 857  
4631 TCGCCACCAAGTTCCTCAAAACGATTACCTTTGTCGATAACTTTGAAAGG 4680  
857 spAspThrPheTyrMetAsnThrMet...LeuIleSerAsnLysValGln 872  
4681 AGGAGTGGCGTCTTTTCTCAAGCTGGGTGAGGGGTATACCTATANTGG 4730  
873 ArgSerSerLeu..... 876  
4731 TGGTAGCCATGTTTCATCAGGTGCGCTCGTCCCTAGAGGATATCTTAA 4780  
877 .....ProLys..... 878  
4781 CGGCAATTAAAGTACCCAAGCGCTTCGACCCACTGTTTAGTCGAGAAGTAC 4830  
878 ..... 878  
4831 AGATGGGTGGAGCGGTACCATTCACGCTGATCAGGAGGTGCTATCC 4880  
879 .....GlnPr 880  
4881 ATCAGATAACCCCTATCTTCAACGCTCAATCTCGTGGGAAGCAAACTTC. 4929  
880 oAspPheAsnLysTyrAsnThrValGlnGlnLysAlaLysGlnGluTyrL 897  
4930 .....TCGACTAAGTGCAGGAAGGTGTAAAGTTCATGCTCATAAACGTA 4974  
897 euTrpTyrLeuArgCysLys.....MetIleSerAsp... 907  
4975 GCTTCGGGTGACTATTTCTTATGCTTCCGCTTTTCAAAGGACGCGACTT 5024  
908 .....ArgThrThrLe 911  
5025 GCATTTCAGTAAATCCATCGACGAAGGGCGCATCAGTTTGACGTTTCAGG 5074  
911 uArgSerIle..... 914  
5075 CAACTCGCGCGCTCTTGGTGTAGGAGGATGTTGCAGTTAGCCGCGCGC 5124  
914 ..... 914  
5125 GTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCGACCAACACAGAGCCA 5174  
915 IleGluAspHisLeuAsnGlyLeuTyrHisAsn..... 925  
5175 AGGTGCTACCAACAACAAATCACACCAAAATCGGGGGCAAGGCTCTATCTG 5224  
926 .....GlyAsnCysAspLeuProL 932  
5225 AGGGAAGTGGTAGGAAAGTCAAGGGGAGGTGCACATCTCGATATGTTGC 5274  
.....

932 yAsnSer.....Cys 935  
5275 GAACAAGATTACGTTAGGAAGTGTAGTGGCTCAGGCGTGATAATCCAGT 5324  
|||||.....|  
936 PheLeuAspThrLysSerValGlyGly..... 945  
5325 GATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAA 5374  
946 .....GluTrpLeuTyr 949  
5375 CCGGGACCTCTGAAGATGCCGTGGTGAGTACTTGGCTAGAAATATTCGGT 5424  
|||||  
950 AlaLysProLeuLys..... 954  
5425 GGCATTGACAGGACATACAGGGCGTTGCTTATGCGTAGAAATATTCGGT 5474  
:|||||  
955 .....AlaGlyHisCys.TyrGly..... 960  
5475 CACTACCGCGGAGGTGTTCTGTAAGTACTCAATCAAGTTTATGAATCAC 5524  
960 ..... 960  
5525 TACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTC 5574  
|||||.....ValGly.....PheSerLeuA 966  
5575 ACACAAGACGGCTGCTGCTGTGAGAGACCTACCGTACTCATAGCTGA 5624  
|||||  
966 snSerLys.Gly.....LysIleGlyLy 973  
5625 AAAAGGTATCTTTACCACGAGGCAAGATGTCGACGCG..... 5661  
:.....|  
973 sCysGluLeuLeuLysLeuMetTrpAspValAspAlaArgGlyGlnProL 990  
5662 .....GTGGTAGCTTTGGGCGCAC 5679  
990 euGluArgProPheAsnThrLysMetPheGlnPheValLeuLeuAsnAsp 1006  
5680 AATCTCTTGTATCCAGCATATACGTTTCCAGCATGCCATTAATTT 5729  
|||||  
1007 LeuThrPheLeuMetAsnGluMetIleIlePheArg.....AsnLe 1020  
5730 GATAGTGTCACGAAAGTCGCTCGATGCGCATGGTGGCGAATCGTTTA 5779  
|.....|  
1020 uGlnAspThrLeuGlnArgLys.....GluArgThrL 1031  
5780 AGTCCTTCGAATATAAGTCTAATAATGCTCCCGCAGGTGGCGTAAGACG 5829  
|.....|  
1031 yscGlnAlaSerIleValLeuLysAspGlyValProGlyCysGlyLysSer 1047  
5830 ACGACCTTAGTGGACCAATTCGTTAAGTCACCAATAGCACACCCACCAT 5879  
|||.....|  
1048 ThrTrpIleLeuAsn.....AsnAlaAsnTyrIleLysAspVa 1060  
5880 TAGCGGTATGTGGCA.....AGTTCTGAGGACATAAATATGGCGGTGA 5923  
:.....|  
1060 ValIleSerValGlyLysGluAlaLysGluAspLeuLysGluLysPheM 1077  
5924 AGAAGAGATCCGAATTTGGAAGT...CTCAACAGTGTACCACAGTT 5970  
|.....|  
1077 etLysLysTyrLysCysThrGluSerGluLeuGlyArgIleArgThrVal 1093  
5971 ACTCCAGGCTGTAACCTTTATCGTCAGGGAATATATAAAGGCTTTT 6020  
:|||||  
1094 AspSer.....TyrLeuMetHisAspCysGlyLysLysLeuAr 1106  
6021 GGTGGATGAGTGCAC.....ATGATGCATCAGGCTTAC 6055  
:|||||  
1106 gAlaThrThrValHisPheAspGluAlaLeuMetThrHisAlaGlyAlav 1123  
6056 TACAACATGGCGCTTCGCGAACCGCGCTCGGAGGCGCTCTTTTGGGA 6105  
:.....|  
1123 altyrPheCysAlaAspLeuLeuGlyAlaArgLysValIleCysGlnGly 1139

6106 GACATAAATCAGATACATTATCAAAACAGGAGAGAGTGTATTAGGATGGA 6155  
|||.....|  
1140 AspSerGlnGlnIleProPheValAsnArgValGluSerIleLysLeuG 1156  
6156 TTGTGCTGTTTGTTCCTCAAGAGGAAAGCGTTGTATACACTTCTTAAT 6205  
|||.....|  
1156 nPheAlaLysLeuValIleAspLysThrAspLeuLe.....ArgMetT 1171  
6206 CGTACAGGTCTCCGTTAGATGTTTGTCTACTTGTCTCTCAATGACCGTA 6255  
|||.....|  
1171 hrTyrArgSerProIleAspValAlaHisTyrLeuAsn..... 1183  
6256 AGGGACCGGAAAGTGTACCTCTGAAAGGTCTTACCGGTAG..... 6300  
|||.....|  
1184 .....TyrLysSerPheTyrThrGlyGlyArgIleTh 1194  
6301 .....GACAAACCACTAGTAAAGTCGCTGCC.....A 6328  
1194 rThrLysAsnGluValValArgSerMetSerValValGlyProArgAsnV 1211  
6329 AAAGGCCAATTGGAACCACTGATCAGCTGAAATAAAACGCTGACGTG 6378  
|||||.....|  
1211 alArgProMetThrSerValTyrSerValProGlyValGln 1227  
6379 TACTTGTGATGATCCAGTTGGAGAGTCGGATATGAAGAGTCTGTGAA 6428  
|||||.....|  
1228 TyrLeuThrPheThrGlnThrGluLysAspLeuPheLysAlaLeuAr 1244  
6429 GGGAAAGGAAAGAACACCACTGATGACATGCATGAAAGCACAGGAA 6478  
1244 gSerLysGly...HisValAsnValAsnThrValHisGluThrGlnGlyL 1260  
6479 AAACATTACGTGATGTGTTATTAGTACGAGAAAGACCGCTGACTCC 6528  
|||||.....|  
1260 yThrPheAspValIleLeuValArgLeuLysThrThrGluAsnGlu 1276  
5529 CTATTC.....ACTAAACACCGCATATCTTGTGTGTGTC 6566  
:.....|  
1277 ELeTyrProGlyGlyArgAsnSerLysProTyrThrIleValGlyLeuTh 1293  
6567 GACACACACGCTCAGTGGT...TATGCCGCTCTCAGCTCAAGTTG 6612  
1293 rArgHisArgSerLeuValTyrThrAlaIleGluAspArgLeu 1309  
seq\_name: sp\_virus:Q9WIE3  
seq\_documentation\_block:  
ID Q9WIE3 PRELIMINARY; PRT; 1350 AA.  
AC Q9WIE3;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE REPLICASE.  
OS Chinese wheat mosaic virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.  
OX NCBI\_TaxID=83544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-YANTAI, CHINA;  
RX MEDLINE=9281891; PubMed=10355760;  
RA Diao A., Chen J., Ye R., Zheng T., Yu S., Antoniw J.F., Adams M.J.;  
RT "Complete sequence and genome properties of Chinese wheat mosaic  
virus, a new furovirus from China";  
RL J. Gen. Virol. 80:1141-1145(1999).  
DR EMBL; AJ012005; CAB41770.1; -.  
DR INTERPRO; IPR00086; -.  
DR INTERPRO; IPR002588; -.  
DR PFAM; PF01443; Vmethylcasel; 1.  
DR PFAM; PF01660; Vmethyltransf; 1.  
SQ SEQUENCE 1350 AA; 153005 MW; F8D84D51BD42681F CRC64;



```
2655 TTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGTGCCTTTGATA 2704
503 .....GlyThrAspAlaIleGlnPheHisAsp..... 511
2705 TTTTCAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGC CGCTC
512 .....Val 512
2755 TTCTCTAGGCTATCGTGGGAAGATTCAAATTAAGTCATGAAGCAATGAA 2804
513 ProLeuTyrAlaGluValThrAspArgValLysLeuTrpLysThrHisAl 529
2805 GACAGAAGATGGAACCCCTGCCAATTACTGAAGATTCTGTATATCCGT 2854
529 aProAsnGlnGlyPheValLeuAspMet...GluAspLeuAspValLysI 545
2855 TCATAATGGGGAACGTTTCTTAACGTCCTCACTGTACGAGGGCAGGTCTTCT 2904
545 LeLysMetHisGluValSer..... 551
2905 GCGGTTTCGAACCGACCGTGGTTTCGAGTGTTCCTAAGGGTTTGGTAGC 2954
551 ..... 551
2955 TCCTGGGGCTGCGACGAGGCCCTTTTCTGGCATTACGTCGTTCTTTCCA 3004
551 ..... 551
3005 CAGGTTCACTATTCTACGACCGCGTTTAACTGAAGATGAAGCGTTGAT 3054
552 .....GluArgGluLysArgAsp 557
3055 GCTCTGTGCGGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGA 3104
557 ..... 557
3105 GACGTCGCGGTAGCTGTGAGCAAGCTGTAGCTGGACGAAGAATTTT 3154
558 .ValSerArgCysIleValSerGlyLysLeu...GlyGluLeuGlnIleH 573
3155 GGAGTGAAGTTCCTTAAATGACTTCCACCACCTTTCGTATTGCGGAATAAG 3204
573 IsSerGluSerSerAsnAsnGluGluGlnIle.....Asn..G 585
3205 GTGCTTATCGGGATATTCGTGGCGCTTTCGGGTGCGGCGCCCAATTCATG 3254
585 LyAspTyrArgAsp..... 589
3255 GAAGTATAGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTA 3304
589 ..... 589
3305 GTTACGAACCTCTAAGCTCGTTAAAGTTCAACAGCCGCGGTGTTTACGC 3354
590 .....SerArgArgArgThr.. 595
3355 GGTTTAACTCTAGCACAGTATCCGGTGGATCTTTAGTCGTCGGAAGAGG 3404
595 ..... 595
3405 GTTTTCGTCGGCGGTGACCGTCACTAGGGGGACCGTAGCTAAACGTCAG 3454
595 ..... 595
3455 TCCCTTTAGCGTTGCTATCGTTTCTTACCTCATACGCCATTTCCGGCTGC 3504
595 ..... 595
3505 AGTATGTTAGGCAATTGGGCACATGCTCTTCCACGGCACTTAATGTTTT 3554
595 ..... 595
3555 CTTTGGTTTAGGACATTCGTTGGGGCGAGGGCTACGCGGAATACTGGA 3604
```

```
596 .....PheGluAspLeuLeuGluGlyGluValAlaAlaThrAsnPheLeuAs 610
3605 AGTTTGGAGCTTCTCCAATAATTGGTGGCGCTGTTCCCGAGGT..... 3647
610 pAspTrp.....CysGluLysAsnAspHisLsp 619
3648 .TGTTTGGCGAGGAAGAGTGTCACTGCTTGTACTGCTTATACGCT 3695
619 heAsnPheSerArgAspAspAlaGln.....SerLysTyrAla 631
3696 AGGGTATCTTTGATCATAAGGGCTTCTTAACGACACACCATACCTCAAC 3745
632 TrpGlyLeu..LysLeuLeuLysGlyIleTrpGluPheLeuLeuProPom 648
3746 TTGCTTACGTCGCCACCGGTAGAGGGGAGGAATGTGTAC.....GATCAG 3789
648 etAspPheAlaPro.....ValTyrValAspAspGlu 658
3790 ACGTTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGTCCATC 3839
659 GlnAlaArg...LeuArgMetValArgValAsnGluArgArgAsnThrSe 674
3840 TGGGACTCAGCATCAACCGGTTCCCGGTGAGGATACGATGGATCCACTT 3889
674 rasPThr.....GlyAlaAspGlyAspValAlaAlaC 685
3890 CTAGTGTCTCAAGC.....TATGATGCTT 3912
685 ysGluValAlaAlaAspPheSerLysAlaMetaspThrLeuValaspVal 701
3913 GTCACAATGTG.....COCGACGTGGGATTAGCACCAACGGGGAAGT 3956
702 VallysLysMetaspGlnLysLysLeuGlyLeu..... 712
3957 TACTGGTGAAGAGAGACCCATTCACCTCGACGCTGCAATACACTTATG 4006
713 .....V 713
4007 TCGAGGAAGAGGTTCGCCCTCTGCAGCTGTGGCGGAAGACAAAGGTGAT 4056
713 alGluLysValLysAlaSerAlaValAlaValGluLeuSerLysThr 729
4057 CCCTCGGGTTCGTGTCACCGTGTGCTTGTGTTGTTGAAAGTGTGAA 4106
730 LysProValThrGlyAsnSerCysAlaValAspLeuTrpAlaAspPheG 746
4107 AAAGGTGTCGACGATGCTCTTTCACCAACAGCTAGTGGGGGAACGGCTC 4156
746 uLysAsnLeuAspAsp.....AspAspGluProGlyCysGlyI 759
4157 GTGAGGTTCAGGTGGACGCAAGGTTGCTCCAGAAAGCGTCGTCGCT 4206
759 leSerValValThrLysGlyLysGluIleCysGlnAspAsnVal..... 773
4207 GAGCGCCGACACAAAGAAAGGGAAGAGCTGCAGATGGTAAACACAGCACA 4256
774 ..GlyProVal.....LeuLeuCysGlySerSerSerCy 784
4257 AACCGGCGTCAACGAGCGCACAGGGAGCGCAGTACAGTCCAGTCTTGTGA 4306
784 sSerSerValSerGluValGluLysGlu..... 793
4307 GTTCGCCACAGGCTGATATTCCAAAGGTCAACCCAGTCCGAGGTACATGCT 4356
794 .....ThrAspValValSerValThrAsp..... 801
4357 CAGAAAGAGTGAACAAAGAAAGTACCATTGCGGACTGTTTCGGGCGCAC 4406
802 .....SerGlyAsnSe 805
4407 GCCAATCGTCGATGAGAACCCGCCCAAGCTGTACAGCTCGTGGTGTGA 4456
:|||||
```





*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 42.21 Seconds  
(without alignments)  
97.211 Million cell updates/sec

Title: US-09-301-906-5  
Perfect score: 605  
Sequence: 1 VGSVSNLRGDKKVLMEAR.....EPAQSVPOQSLEKALTFG 120

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 3419795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	605	100.0	120	21 Y58143	GLRAV-3 polyprotei
2	605	100.0	2237	21 Y58148	GLRAV-3 polyprotei
3	80.5	13.3	3010	15 R53417	Blood transniscibl
4	78	12.9	483	14 R43893	elav. Drosophila
5	78	12.9	485	19 W60577	Drosophila neuron-
6	74	12.2	1301	18 W29115	FMK2P protein asso
7	71.5	11.8	3011	13 R22154	NANBV Hutch c59 is
8	71	11.7	285	19 W98314	H. pylori GHPO 894
9	70	11.6	1780	19 W53863	Human gravin polyp
10	69.5	11.5	380	12 R10748	Non-A non-B hepati
11	69.5	11.5	1061	20 W87504	Human N-methyl-D-a
12	69.5	11.5	1212	20 W87503	Human N-methyl-D-a

13	69.5	11.5	1219	18 W25763	Amino acid sequenc
14	69.5	11.5	3011	15 R66995	Hepatitis C virus
15	69	11.4	509	17 W05399	Human clone 65 pro
16	69	11.4	730	20 Y43385	S. pneumoniae PspC
17	69	11.4	1237	21 Y81609	Streptococcus pneu
18	68.5	11.3	248	18 W20433	H. pylori secreted
19	68.5	11.3	248	18 W20610	H. pylori secreted
20	68.5	11.3	449	19 W37129	Hepatitis C virus
21	68.5	11.3	1528	18 W33363	Human multidrug re
22	68.5	11.3	3011	14 R34468	Encoded by full-le
23	68	11.2	328	19 W36984	E. coli k1pA prote
24	68	11.2	390	17 W05283	Truncated human ta
25	68	11.2	441	15 R58810	Human tau protein.
26	68	11.2	441	17 W05282	Human tau protein.
27	68	11.2	441	18 W34856	Human tau protein.
28	68	11.2	441	21 Y81386	Human paired helic
29	68	11.2	3969	15 R52971	Product of the cDN
30	67.5	11.2	391	15 R50080	NANBH virus antige
31	67.5	11.2	1527	20 Y43543	A human MPR-relate
32	67.5	11.2	2783	13 R23962	AFP-1. Homo sapie
33	67.5	11.2	2783	13 R23963	AFP-1 (Ala 2460 Va
34	67	11.1	554	19 Y20763	Human neurofilamen
35	67	11.1	1400	15 R44514	MLL amino acid seq
36	67	11.1	1400	17 R92705	MLL cDNA clone 14-
37	67	11.1	3910	14 R38470	ALL-1 protein. Ho
38	67	11.1	3910	16 R66462	ALL-1 (acute lymph
39	66.5	11.0	238	18 W14587	Streptococcus pneu
40	66.5	11.0	250	20 Y41519	Fragment of human
41	66.5	11.0	346	21 Y58628	Protein regulating
42	66.5	11.0	411	17 R90934	HCV NS5 domain ant
43	66.5	11.0	479	20 Y03184	Topoisomerase II b
44	66.5	11.0	499	19 W67010	HCV non-structural
45	66.5	11.0	504	12 R14553	Chimpanzee HCV clo

## ALIGNMENTS

RESULT 1

- Y58143.  
ID Y58143 standard; Protein; 120 AA.  
XX  
AC Y58143;  
XX  
DT 07-MAR-2000 (first entry)  
XX  
DE GLRAV-3 polyprotein proteinase domain.  
XX  
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;  
KW viral disease; yield loss; sugar content; inhibition; infection;  
KW replication; polyprotein; domain; proteinase; methyltransferase;  
KW helicase; RNA-dependent; RNA polymerase; untranslated region;  
KW transgenic plant; component; resistant; truncation; deletion; antisense;  
KW expression; detection; antibody.  
XX  
OS Grapevine leafroll-associated virus 3.  
XX  
PN W09955880-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 29-APR-1999; 99WO-US09307.  
XX  
PR 29-APR-1999; 98US-0083404.  
XX  
PA (CORR ) CORNELL RES FOUND INC.  
XX  
PI Gonsalves D, Ling K;  
XX  
DR WPI; 2000-062035/05.  
XX  
DR N-PSDB; Z49203.  
XX  
PT Newly isolated grapevine leafroll virus protein or polypeptide useful

for producing transgenic plants conferring viral disease resistance -

Claim 4; Fig 6; 84pp; English.

This sequence represents the proteinase domain of the polyprotein (Y58148) from the grapevine leafroll-associated virus 3 (GLRAV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, which may serve as targets for the inhibition of viral infection or replication. These proteins include the 242-248 kD polyprotein (Y58148), encoded by open reading frame (ORF) 1a, and which comprises a proteinase domain (Y58143), a methyltransferase domain (Y58144) and a helicase domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF 1b; and a protein of unspecified function (Y58147), encoded by ORF 1c. Nucleotides encoding these proteins, or fragments thereof, and the 5' and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are useful for the generation of transgenic plants and plant components. Such transgenic plants may be resistant to viral disease, for example, this property being conferred on the plants via the use of nucleotides encoding truncated or internally deleted proteins, or via the use of antisense nucleotides to inhibit viral gene expression. The nucleotides may additionally be used in the detection of viral nucleic acids in a tissue sample. The proteins can be used to generate antibodies which may be used to detect GLRAV-3 in plant samples. The isolation of GLRAV-3 facilitates the production of agents that reduce the risk of infection or damage by the virus in vineyards.

Sequence 120 AA;

Query Match 100.0%; Score 605; DB 21; Length 120;  
Best Local Similarity 100.0%; Pred No. 3, 7e-61;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 VSGSVSALRGDGKKVLMEARTSATSDVSDFDVFESVSNALLVVHYHRVVPYPVKRE 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1 vsgsvsalrgdgkkvlmeartsatdsdvdvfeavsnallvvhyhrvvpypvkrc 60  
  
QY 61 QPKPAVKQDBQKPRQASHWAVKPTAGVHVHPLPKQEALEPAQSVPQOSLEEKAALTFC 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 61 qp xpavkqdesqpkrgashwawkptagvhvplpkpgealepaqsypqgsleekaaltfc 120

RESULT 2

Y58148 ID Y58148 standard; Protein; 2237 AA.  
XX Y58148;  
XX  
DT 07-MAR-2000 (first entry)  
XX  
XX GLRAV-3 polyprotein.  
XX  
XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;  
KW viral disease; yield loss; sugar content; inhibition; infection;  
KW replication; polyprotein; domain; proteinase; methyltransferase;  
KW helicase; RNA-dependent; RNA polymerase; untranslated region;  
KW transgenic plant; component; resistant; truncation; deletion; antisense;  
KW expression; detection; antibody.  
XX  
OS Grapevine leafroll-associated virus 3.  
XX  
XX Key Location/Qualifiers  
FH 85..204  
FT Domain /note= "Proteinase domain (Y58143)"  
FT 460..731  
FT Domain /note= "Methyltransferase domain (Y58144)"  
FT 1922..2212  
FT Domain /note= "Helicase domain (Y58145)"  
XX  
XX WO9955880-A1.  
XX





```

Db 13 gggvdtqqlmqsaasaaavaatnaaaapvqnaa-----avaaaaqlggqvgqailqv 66
QY 70 EOKPKRQASHWAVPTAVGVHVPLPKKQEALEPAQSVPOQSLEEKAAAL 117
   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 qggqgtqgq---avaaaaavtqqlggqvgqavvaqgavvqggqgqaaav 110

RESULT 6
W29115
ID W29115 standard; Protein: 1301 AA.
AC W29115;
XX
XX
DT 15-JAN-1998 (first entry)
XX
XX FMR2p protein associated with FRAXE mental retardation.
XX
KW FMR2 gene: FMR2P; FRAXE; rate folate-sensitive fragile site;
KW X-linked mental retardation; diagnosis; therapy; antibody.
XX
OS Homo sapiens.
XX
PN WO9723610-A1.
XX
PD 03-JUL-1997.
XX
PF 20-DEC-1996; 96WO-AU00825.
XX
PR 22-DEC-1995; 95AU-0007366.
XX
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI Gecz J, Mulley JC;
XX
DR WPI: 1997-351051/32.
DR N-PSDB; T85728.
XX
XX DNA containing gene associated with FRAXE mental retardation -
PT useful for diagnosis and therapy of FRAXE mental retardation
XX
PS Claim 36; Fig 2: 39pp; English.
XX
CC This polypeptide comprises FMR2P protein encoded by the human FMR2
CC gene (see T85728). Absence of FMR2P appears to be the molecular
CC basis for FRAXE (a rare folate-sensitive fragile site), a
CC non-specific, mild, X-linked form of mental retardation. This can
CC be caused either by CCG expansion within the 5'-untranslated region
CC of the FMR2 gene (see T85728) or by deletion of coding sequences.
CC An isolated FMR2 CDNA sequence can be used in a claimed recombinant
CC expression system to produce a protein product in transformed host
CC cells. This can be used to raise polyclonal or monoclonal
CC antibodies that form the basis of an immunocytochemical test for
CC the presence of FMR2P. The protein encoded by the normal
CC chromosome can also be used to treat FRAXE mental retardation by
CC replacing, repairing or compensating for the FRAXE allele.
XX
SQ Sequence 1301 AA;

Query Match 12.2%; Score 74; DB 18; Length 1301;
Best Local Similarity 32.4%; Pred. No. 8.5;
Matches 23; Conservative 12; Mismatches 28; Indels 8; Gaps 3;

QY 50 RVVPYVPKREQPKPAVKODEQPKRQASHWAV-KPTAVGVHVPLPKKQEALEPAQSVPO 108
   || : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 802 rvpghsslnhaapakdhketatpkrtqrtavtavekpkg-----krkhkpievaeakipe 856
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 109 --QSLEEKAAAL 117
   | | | | |
Db 857 kkgrileeatti 867

```

```

RESULT 7
R22154
ID R22154 standard; Protein: 3011 AA.
XX
AC R22154;
XX
XX 07-JUL-1992 (first entry)
XX
XX NANBV Hutch c59 isolate genome.
XX
XX Hepatitis C virus; non-A non-B virus; HCV-Hc59; antibodies;
KW vaccine; assay; detection.
XX
XX NANBV Hutch c59 isolate.
XX
XX
FH Key
FT Region
FT 1..326
FT /label= NANBV_structural_proteins
FT /note= "Seq No 1 (R22137)".
FT Misc-difference 3
FT /label= Asn
FT /note= "or Ile according to Seq No 1 (R22137), see CC"
FT Misc-difference 321
FT /label= Asn
FT /note= "or Asp according to Seq No 1 (R22137), see CC"
FT Region
FT 1..120
FT /label= capsid
FT Region
FT 121..326
FT /label= envelope_protein
FT Peptide
FT 1..20
FT /label= pref._capsid_antigen
FT Peptide
FT 21..40
FT /label= pref._capsid_antigen
FT Peptide
FT 2..40
FT /label= pref._capsid_antigen
FT Peptide
FT 1..74
FT /label= pref._capsid_antigen
FT Peptide
FT 69..120
FT /label= pref._capsid_antigen
FT Peptide
FT 121..176
FT /label= pref._envelope_antigen
FT Region
FT 386..411
FT /label= V_variable_region
FT Region
FT 246..275
FT /label= V1_variable_region
FT Region
FT 456..482
FT /label= V2_variable_region
FT Region
FT 2356..2379
FT /label= V3_variable_region
FT Peptide
FT 391..404
FT /label= pref._V_fragment
FT Peptide
FT 246..256
FT /label= pref._V1_fragment
FT Peptide
FT 461..466
FT /label= pref._V2_fragment
FT Peptide
FT 473..482
FT /label= pref._V2_fragment
XX
XX WO9203458-A.
XX
XX 05-MAR-1992.
XX
XX 23-AUG-1991; 91WO-US06037.
XX
XX 21-NOV-1990; 90US-0616369.
XX 25-AUG-1990; 90US-0573643.
XX
XX (NYBL-) NEW YORK BLOO DCENT.
XX (PHAA ) PHARMACIA GENETIC ENG INC.
XX
XX Zebedee S, Inchauspe G, Nasofe MS, Prince AM;
XX WPI: 1992-096821/12.
DR

```









pb 203 asvspib

203 qsvspibhqqggyvenlkagalcswtakkdnhlnfskhditylenqenwwfney



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 27.64 seconds  
(without alignments)  
77.961 Million cell updates/sec

Title: US-09-301-906-5  
Perfect score: 605  
Sequence: 1 VSGSVSALRGDKKVLMEAR.....EPAGSVPQSLKKAALTFG 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	78	12.9	485	1 US-07-881-075-1	Sequence 1, Appli
2	78	12.9	485	1 US-08-120-827-1	Sequence 1, Appli
3	78	12.9	485	1 US-08-478-675-1	Sequence 1, Appli
4	70	11.6	1780	1 US-08-769-309A-5	Sequence 5, Appli
5	70	11.6	1780	3 US-08-994-570-5	Sequence 5, Appli
6	69.5	11.5	1199	1 US-08-041-538-2	Sequence 2, Appli
7	69.5	11.5	1199	1 US-08-463-642-2	Sequence 2, Appli
8	69.5	11.5	1199	1 US-08-455-602-2	Sequence 2, Appli
9	69.5	11.5	1199	2 US-08-465-157-2	Sequence 2, Appli
10	69.5	11.5	1199	4 PCT-US91-09422-2	Sequence 2, Appli
11	69.5	11.5	1219	2 US-08-687-289A-6	Sequence 6, Appli
12	68	11.2	328	2 US-08-651-818A-2	Sequence 2, Appli
13	67	11.1	1400	1 US-08-080-255-7	Sequence 7, Appli
14	67	11.1	1400	3 US-08-465-713-7	Sequence 7, Appli
15	67	11.1	1400	4 PCT-US93-05857-7	Sequence 7, Appli
16	66.5	11.0	504	1 US-07-853-985A-6	Sequence 6, Appli
17	66.5	11.0	504	1 US-07-681-703B-6	Sequence 6, Appli
18	66.5	11.0	504	1 US-08-184-236-6	Sequence 6, Appli
19	66.5	11.0	504	2 US-08-407-410B-6	Sequence 6, Appli
20	66.5	11.0	504	2 US-08-485-500-6	Sequence 6, Appli
21	66.5	11.0	504	4 PCT-US91-02370-6	Sequence 6, Appli
22	66.5	11.0	504	4 PCT-US94-04174-6	Sequence 6, Appli
23	66.5	11.0	541	4 PCT-US92-06565A-13	Sequence 13, Appli
24	66.5	11.0	872	1 US-08-491-357-3	Sequence 3, Appli
25	66.5	11.0	872	3 US-08-968-633-3	Sequence 3, Appli
26	66.5	11.0	872	3 US-09-196-466-3	Sequence 3, Appli
27	66.5	11.0	872	4 PCT-US96-10823-3	Sequence 3, Appli
28	66.5	11.0	1786	3 US-08-444-818-54	Sequence 54, Appli

29	66.5	11.0	2261	3 US-08-444-818-66	Sequence 66, Appli
30	66.5	11.0	2436	3 US-08-444-818-75	Sequence 75, Appli
31	66.5	11.0	2772	3 US-08-444-818-89	Sequence 89, Appli
32	66.5	11.0	2894	2 US-08-466-975A-23	Sequence 23, Appli
33	66.5	11.0	2894	2 US-08-391-671A-23	Sequence 23, Appli
34	66.5	11.0	2894	3 US-08-467-902A-23	Sequence 23, Appli
35	66.5	11.0	2955	2 US-08-443-260-3	Sequence 3, Appli
36	66.5	11.0	2955	3 US-08-442-805A-3	Sequence 3, Appli
37	66.5	11.0	2955	3 US-08-443-900A-3	Sequence 3, Appli
38	66.5	11.0	2955	3 US-08-444-818-124	Sequence 124, Appli
39	66.5	11.0	2995	3 US-08-444-818-138	Sequence 138, Appli
40	66.5	11.0	3011	1 US-08-440-103-36	Sequence 36, Appli
41	66.5	11.0	3011	1 US-08-440-542-36	Sequence 36, Appli
42	66.5	11.0	3011	1 US-07-910-760-10	Sequence 10, Appli
43	66.5	11.0	3011	1 US-08-440-519-10	Sequence 10, Appli
44	66.5	11.0	3011	1 US-08-231-368-36	Sequence 36, Appli
45	66.5	11.0	3011	1 US-08-440-210-36	Sequence 36, Appli

ALIGNMENTS

RESULT 1  
US-07-881-075-1  
; Sequence 1, Application US/07881075  
; Patent No. 5444149

; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
; ADDRESS: P.C.  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/881,075  
; FILING DATE: 19920511  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5444149man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-154-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)521-4500  
; TELEFAX: (703)486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 485 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-07-881-075-1

Query Match 12.9%; Score 78; DB 1; Length 485;  
Best Local Similarity 25.9%; Pred. No. 0.18;  
Matches 28; Conservative 22; Mismatches 48; Indels 10; Gaps 2;

QY 10 GCGKVLMEARSTSTSDVDFVVFVAVSNALLVHVHVVPVAPVREPKPAVKOD 69





```

Db      993  PHLTAETPTFLAD--SVIPK--GLPPPLPQQQPQQPPPPQQPQQ 1033

RESULT      7
US-08-463-642-2
; Sequence 2, Application US/08463642
; Patent No. 5721107
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,642
; FILING DATE: 03-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,007
; FILING DATE: 18-MAR-1991
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; APPLICATION NUMBER: 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1199 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-642-2

Query Match      11.5%; Score 69.5; DB 1; Length
Best Local Similarity 32.4%; Pred. No. 7.3;
Matches 34; Conservative 11; Mismatches 47; Indels

QY      9  RGCGKKVLMPEARTSTSATSDVSDFD----VFEEVSNALLVVVHHVHVPPY
      :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
Db      938  QSGKGS--LTFSADSTKLTLYNVEEDNTPSAHFSPSSPMVY--HRRG-
      :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
QY      65  AVQDCKPKRQASHNAVPTAGVHVHPLPKQEALEPAQSVPPQ 109
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      993  PHLTAETPTFLAD--SVIPK--GLPPPLPQQQPQQPPPPQQPQQ 1033
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT      8
US-08-455-602-2
; Sequence 2, Application US/08455602
; Patent No. 5747267
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R

```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-687-289A-6

Query Match 11.5%; Score 69.5; DB 2; Length 1219;
Best Local Similarity 32.4%; Pred. No. 7.5;
Matches 34; Conservative 11; Mismatches 47; Indels 13; Gaps 6;

QY 9 RGDGKKVLM EARTSTTSATS DVSDFD---VVF EAVSNALLVYHVRVVPYAPVKRQPKP 64
Db 958 QGSGKS-LTFS DASTKTLNVEEDNTPSAHFSPSPSMVV--HRRGP--PVATTPLP 1012

QY 65 AVKDEQPKRQASHWAVKPTAVGVHPLPKKQEALEPAQSVPOQ 109
Db 1013 PHLTAETPLFLAD--SVIPK--GLPPLPQQQPPQPPQPPQ 1053

RESULT 12
US-08-651-818A-2
; Sequence 2, Application US/08651818A
; Patent No. 594889
; GENERAL INFORMATION:
; APPLICANT: de Boer, Piet A.J.
; APPLICANT: Hale, Cynthia A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING
; TITLE OF INVENTION: ANTIMICROBIALS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651.818A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02249
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant

```

```

; MOLECULE TYPE: protein
; US-08-651-818A-2

Query Match 11.2%; Score 68; DB 2; Length 328;
Best Local Similarity 22.8%; Pred. No. 1.8;
Matches 34; Conservative 11; Mismatches 50; Indels 54; Gaps 4;

QY 6 SALRGDGKKVLM EARTSTTSATS DVSDFDVF EAVSNALLVYHVRV-----VRVHRVNHAPANAQEH EAA R 51
Db 34 SMFDRDRPLKRMKSKRDDSDYDEDEDGSGE-----VRVHRVNHAPANAQEH EAA R 85

QY 52 -----VPYAPVKRQ P-----KPAVKDEQPKRQASHWA 81
Db 86 PSQHQYQPPYASAPRQPVQVQPPPEAQVPPQHAPHAPQVQVQPPQEQPLQQPVSPQ 145

QY 82 VKPTAVGVHVP LPKKQEALEPAQSV--PQ 108
Db 146 VAPAPQPVHSAPFAQQAQFAEPVAAFPQ 174

RESULT 13
US-08-080-255-7
; Sequence 7, Application US/08080255
; Patent No. 5487970
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 19930617
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-080-255-7

Query Match 11.1%; Score 67; DB 1; Length 1400;
Best Local Similarity 23.5%; Pred. No. 18;
Matches 23; Conservative 14; Mismatches 49; Indels 12; Gaps 3;

QY 18 EARTSTTSATS DVSDFDVF EAVSNALLVYHVRVVPYAPVKRQPKPAVKDEQPKRQ 77
Db 440 EKSKTSEKKDSKSSSVVKNVVDSS-----OKPTPSA---REDPAPKKSSEPPPRKPV 490

```





***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 60.61 Seconds  
(without alignments)  
134.435 Million cell updates/sec

Title: US-09-301-906-5  
Perfect score: 605  
Sequence: 1 VGSVSALRGDGKKVLMKAR.....EPQSVPPQSSLEKALTFG 120  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues  
Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_66:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	13.8	282	2 A75473	probable sigma fac
2	80.5	13.3	3010	1 A45573	genome polyprotein
3	79	13.1	1545	2 T26589	hypothetical prote
4	78	12.9	483	2 A33130	elav protein - fru
5	77.5	12.8	330	2 E83084	rod shape-determin
6	76.5	12.6	733	2 A45301	microtubule-associ
7	76.5	12.6	860	2 T23296	hypothetical prote
8	74.5	12.3	2359	2 T03094	A-kinase anchor pr
9	73.5	12.1	209	2 I50223	enhancer binding p
10	73.5	12.1	1020	1 QPHUH	neurofilament trip
11	73.5	12.1	1272	2 T30248	fragile X mental r
12	73	12.1	1110	2 I51116	NF-180 - sea lamp
13	72.5	12.0	307	2 S30432	hypothetical prote
14	72.5	12.0	438	2 S77655	hypothetical prote
15	72.5	12.0	592	1 LILBY	actin-binding prot
16	72.5	12.0	691	2 A25704	synapsin I - rat
17	72.5	12.0	1072	2 A37127	microtubule-associ
18	72	11.9	630	2 S55645	transcription cont
19	71.5	11.8	417	2 T49847	hypothetical prote
20	71.5	11.8	465	2 S46759	hypothetical prote
21	71.5	11.8	511	2 T04814	dihydrolipoamide S
22	71.5	11.8	906	2 JC5963	stable tubule only
23	71.5	11.8	1102	2 JC6316	probable protein k
24	71.5	11.8	3011	1 GNMVCH	genome polyprotein
25	71	11.7	210	2 A25550	histone H1 - sea u
26	71	11.7	285	2 E64687	siderophore-mediat
27	71	11.7	491	2 A32584	Abd-B protein - fr
28	71	11.7	791	2 C82940	hypothetical prote
29	71	11.7	1474	2 T20488	hypothetical prote

30	71	11.7	2649	2 T51023	hypothetical prote
31	70.5	11.7	455	2 T38275	hypothetical prote
32	70.5	11.7	1011	2 T29806	myosin-binding pro
33	70	11.6	537	2 A46611	neurofilament trip
34	70	11.6	606	2 A43427	nucleolar phosphop
35	70	11.6	699	2 I38073	probable non-ribos
36	70	11.6	1317	2 B83346	gravin - human
37	70	11.6	1684	2 JW0057	histone H1-I - Vol
38	69.5	11.5	261	2 JN0747	hypothetical prote
39	69.5	11.5	270	2 H83619	hypothetical prote
40	69.5	11.5	617	2 T08565	G protein-coupled
41	69.5	11.5	1199	2 A41939	DNA topoisomerase
42	69.5	11.5	1462	2 T05819	genome polyprotein
43	69.5	11.5	3011	1 S40770	elastic titin - hu
44	69.5	11.5	7962	2 I38346	cell division prot
45	69	11.4	497	1 CERCIFY	

ALIGNMENTS

RESULT 1  
A75473  
probable sigma factor - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: A75473  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: A75473  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-282 <WHI>  
A:Cross-references: GB:AE001935; GB:AE000513; NID:g6458517; PIDN:AAF10382.1; PID:g645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0804  
A:Map position: 1

Query Match 13.8%; Score 83.5; DB 2; Length 282;  
Best Local Similarity 26.6%; Pred. No. 0.76;  
Matches 33; Conservative 28; Mismatches 40; Indels 23; Gaps 7;

QY 4 SVSALRGDGKKVLMKARTSTA-----TSDVSDFDVVFEAVSNALLVWVHYRVVPYAPVK 58  
DB 171 AISRLPGLAQEALAKAAVPAPAPPAEPTAEAAE-DDLFPTLSS-----LAATAPVT 220  
QY 59 RQPP-KPAVKQDQKPKROASHWAVKPTA---VGHVPLPKQ-EALEPAQSVPPQSSLE 112  
DB 221 RVEWPDPRRPQREPEETAE--AVLPEASEATLPVSEPLNPAEPTAQSPETAP 278  
QY 113 EKAA 116  
DB 279 EVQA 282

RESULT 2  
A45573  
genome polyprotein - hepatitis C virus (strain JT)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT) (nonstru  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 16-Jun-2000  
C:Accession: A45573  
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata  
Virus Res. 23, 39-53, 1992  
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier  
A:Reference number: A45573; MUID:92295714

A:Accession: A45573  
A>Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: GB:D1168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613  
A:Experimental source: HCV-JT  
A>Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)  
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmembrane  
F:2-115/Product: capsid protein C #status predicted <GPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepacivirin #status predicted <NS3>  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1230-1237/Region: nucleotide-binding motif A (p-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 13.3% Score 80.5; DB 1; Length 3010;  
Best Local Similarity 21.1%; Pred. No. 22;  
Matches 26; Conservative 24; Mismatches 64; Indels 9; Gaps 3;

Qy 1 VGSVSALRGDGKVKVMEARTSTATSDVDFDVFFAVSNALLVVHYRVVPAPV-KR 59  
:  
Db 2237 MGGNITRVESKNVILDSFDPLRAEEDEREVSVAAILRK-----KKFPALPIWAR 2290

Qy 60 EQPPAKVDGRPKP--RQASHWAVKPTAVGVHVPLPKQEALEPAQSVPQOSLEEKAAL 117  
:  
Db 2291 PDYNPPLLSEKWSFDYVPPAVHGCPDLPPTGPIPPPRKKRTVTLTSTVSSALAEATK 2350

Qy 118 TFG 120  
|||  
Db 2351 TFG 2353

RESULT 3  
T26589  
hypothetical protein Y32B12B.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26589  
R:Gardner, A.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z20239  
A:Accession: T26589  
A>Status: preliminary; translated from GB/EMBL/DDDBJ  
A:Molecule type: DNA  
A:Residues: 1-1545 <ML>  
A:Cross-references: EMBL:AL031632; PIDN:CAA21009.1; GSPDB:GN00023; CESP:Y32B12B.4  
A:Experimental source: clone Y32B12B  
C:Genetics:  
A:Gene: CESP:Y32B12B.4  
A:Map position: 5  
A:Introns: 49/2; 311/2; 879/2; 1116/2; 1257/3; 1404/2; 1545/2

Query Match 13.1% Score 79; DB 2; Length 1545;  
Best Local Similarity 26.6%; Pred. No. 14;  
Matches 29; Conservative 15; Mismatches 41; Indels 24; Gaps 4;

Qy 11 DGKVLMEARTSTATSDVDFDVFFAVSNALLVVHYRVVPAPVKKREQPKAVKODE 70  
:  
Db 1323 DAKTVAPKRSSMSRSF--SLPAVIAQPADA-----PEAHPKRIKTEPLVAPPQ 1370

Qy 71 QKPRQRASHWAVKPTAVGVHVPLPKQ---EALEPAQSVPQOSLEEKAA 116  
:  
Db 1371 EKP-----SYSEVRDIAPCLKMETIVRLVPLEPADLSEKIPETA 1410

```

RESULT      9
I50223
enhancer: binding protein - chicken (fragment)
C: Species: Gallus gallus (chicken)
C: Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C: Accession: I50223
R: Ruddell, A.; Bowers, W. J.
J. Virol. 66: 6578-6586, 1992

```



J. Bacteriol. 175, 37-52, 1993  
A:Title: Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus*  
A:Reference number: S30400; MUID:93106972  
A:Accession: S30432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <WUX>  
A:Cross-references: EMBL:X54107  
C:Genetics:  
A:Genome: plasmid pSCL

Query Match 12.0%; Score 72.5; DB 2; Length 307;  
Best Local Similarity 34.2%; Pred. No. 9.8;  
Matches 25; Conservative 8; Mismatches 33; Indels 7; Gaps 2;

QY 51 VVPYAPVKREQPKPAVKQDEQKPKROASHWAVKPTAVGVHVPKPKQEALEP-----AQ 104  
DB 181 VCPDAGGEREAAAPGAGVAEQSAERIAARWAERRAAAGV-VTSPGRVAFAFAEAGWLLSQ 239  
QY 105 SVPOQSLEKAAL 117  
DB 240 GLPAQRLEAAAV 252

RESULT 14  
S77655  
hypothetical protein o438 - *Mycobacterium leprae*  
C:Species: *Mycobacterium leprae*  
C:Date: 11-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 22-Oct-1999  
C:Accession: S77655; S49517  
R:Fsihi, H.; Cole, S.T.  
Mol. Microbiol. 16, 909-919, 1995  
A:Title: The *Mycobacterium leprae* genome: systematic sequence analysis identifies key ca  
A:Reference number: S77652; MUID:96059637  
A:Accession: S77655  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-438 <FSI>  
A:Cross-references: EMBL:Z46257; NID:9559905; PIDN:CAA86359.1; PID:g559908  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994  
C:Genetics:  
A:Start codon: GTG

Query Match 12.0%; Score 72.5; DB 2; Length 438;  
Best Local Similarity 25.5%; Pred. No. 15;  
Matches 27; Conservative 15; Mismatches 39; Indels 25; Gaps 5;

QY 16 LMEARTSTSDVDFVVFVAVSNALLVHYHRV---VPYAPVKREQPKPAVKQDEQ- 71  
DB 321 LLPVETHTRATPHPA-----HPHRLRLTTPYLHPASDTPYPACRQRSA 364  
QY 72 -KPKROASHWAVKPTAVGVHVPKPKQEA-LEPAQSVPOQSLEKA 115  
DB 365 LAPKPAQSHLTLPPTRSQ---SLPQEQTRLPPGRTVNEETREVA 407

RESULT 15  
LUBY  
actin-binding protein ABP1 - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YCR088w  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 31-Mar-1991 #sequence\_revision 31-Dec-1993 #text\_change 05-Nov-1999  
C:Accession: S19503; S19767; S07608  
R:Dusterhoft, A.; Erdmann, D.; Hegmann, J.; Philippsen, P.; Schweitzer, B.; Spiegelberg  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19351  
A:Accession: S19503  
A:Molecule type: DNA  
A:Residues: 1-476 <DUS>  
A:Cross-references: EMBL:X59720; GSPDB:GNO00003; MIPS:YCR088w  
R:Frontali, L.; Grisanti, P.

submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19504

A:Accession: S19767  
A:Molecule type: DNA  
A:Residues: 14-592 <FRO>  
A:Cross-references: EMBL:X59720; GSPDB:GNO00003; MIPS:YCR088w  
R:Drubin, D.G.; Mulholland, J.; Zhu, Z.; Botstein, D.  
Nature 343, 288-290, 1990  
A:Title: Homology of a yeast actin-binding protein to signal transduction proteins an  
A:Reference number: S07608; MUID:90136906

A:Accession: S07608  
A:Molecule type: DNA  
A:Residues: 1-57, S', 59-311, I', 313-592 <DRU>  
A:Cross-references: EMBL:X51780; NID:93321; PIDN:CAA36075.1; PID:g3322  
C:Genetics:  
A:Gene: SGD:ABP1; MIPS:YCR088w  
A:Cross-references: SGD:S0000684; MIPS:YCR088w  
A:Map position: 3R  
C:Superfamily: actin-binding protein ABP1; SH3 homology  
C:Keywords: actin binding; duplication  
F:158-180, 378-401/Region: duplication  
F:539-588/Domain: SH3 homology <SH3>

Query Match 12.0%; Score 72.5; DB 1; Length 592;  
Best Local Similarity 31.1%; Pred. No. 21;  
Matches 19; Conservative 14; Mismatches 11; Indels 17; Gaps 3;

QY 55 APVKREQPKPAVKQDEQKPKROASHWAVKPTAVGVHVPKPKQEALEPAQSVPOQSLEEK 114  
DB 460 APVQKEEP-----EQEELAPSLPSRN-----SIPAPKQEEAPEQA---PEEIEEE 502  
QY 115 A 115  
DB 503 A 503

Search completed: January 31, 2001, 15:27:32  
Job time: 101 sec

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:27:45 ; Search time 30.65 Seconds

(without alignments)

126.437 Million cell updates/sec

Title: US-09-301-906-5

Perfect score: 605

Sequence: 1 VSGSVSALRGDKKVLMEAR.....EPAQSVPOQSLEKKAALTFG 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	80.5	13.3	3010	1	POLG_HCVJT
2	78	12.9	483	1	ELAV_DROME
3	74	12.2	1311	1	FMR2_HUMAN
4	73.5	12.1	1020	1	NFM_HUMAN
5	72.5	12.0	592	1	ABPL_YEAST
6	72.5	12.0	827	1	MOF_DROME
7	72.5	12.0	1072	1	MAP4_BOVIN
8	71.5	11.8	465	1	YHU6_YEAST
9	71.5	11.8	3011	1	POLG_HCVH
10	71	11.7	205	1	HIE_STRPU
11	71	11.7	210	1	HI_LYTP1
12	71	11.7	285	1	TONB_HELPY
13	70	11.6	232	1	HIB_CHITE
14	70	11.6	537	1	MYPH_CHICK
15	70	11.6	1242	1	KPBL_RAT
16	70	11.6	1781	1	AKAC_HUMAN
17	69.5	11.5	260	1	HIL_VOLCA
18	69.5	11.5	1199	1	MGR1_RAT
19	69.5	11.5	1462	1	TFP2_PEA
20	69	11.4	497	1	FTSY_ECOLI
21	68.5	11.3	188	1	CUP8_DROME
22	68.5	11.3	419	1	FAAA_RAT
23	68.5	11.3	432	1	TAU_RAT
24	68.5	11.3	599	1	HMW3_MYCGE
25	68.5	11.3	845	1	NEM_RAT
26	68.5	11.3	1527	1	MRP3_HUMAN
27	68	11.2	328	1	ZIPA_ECOLI
28	68	11.2	441	1	TAU_HUMAN
29	68	11.2	1008	1	GALY_KLULA
30	67.5	11.2	213	1	H13_RABIT
31	67.5	11.2	244	1	TONB_VIBCH
32	67.5	11.2	704	1	NP14_RAT
33	67.5	11.2	882	1	IF2_BORBU

34	67.5	11.2	1612	1	TP2B_CHILO
35	67	11.1	270	1	TONB_HAEIN
36	67	11.1	407	1	Z174_HUMAN
37	67	11.1	842	1	IF2_TREPA
38	67	11.1	915	1	NFM_HUMAN
39	67	11.1	2468	1	MAPB_HUMAN
40	67	11.1	3969	1	HRX_HUMAN
41	66.5	11.0	741	1	IF2_HACST
42	66.5	11.0	784	1	IF2_ENTFC
43	66.5	11.0	1626	1	TP2B_HUMAN
44	66.5	11.0	3011	1	POLG_HCV1
45	66	10.9	187	1	PABA_ECOLI

## ALIGNMENTS

RESULT 1

POLG\_HCVJT

ID POLG\_HCVJT STANDARD; PRT; 3010 AA.

AC Q00269;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);

ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2

(EC 3.4.22.-); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)

(EC 3.4.22.-); PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);

NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)

(P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate HC-JT) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-92295714; PubMed-1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,

Ra Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;

RT "Molecular cloning of hepatitis C virus genome from a single Japanese

carrier: sequence variation within the same individual and among

infected individuals".

RT Virus Res. 23:39-53(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D11168; BAA01943.1; -

DR PIR; A45573; A45573.

DR HSP; P27958; 1HEI.

DR MEROPS; S29.001; -

DR MEROPS; U39.001; -

DR INTERPRO; IPR000745; -

DR INTERPRO; IPR001490; -

DR INTERPRO; IPR002166; -

DR INTERPRO; IPR002518; -

DR INTERPRO; IPR002519; -

DR INTERPRO; IPR002521; -

DR INTERPRO; IPR002522; -

DR INTERPRO; IPR002531; -

DR INTERPRO; IPR002868; ..  
DR PFAM; PF01560; HCV\_NS1; 1.  
DR PFAM; PF01538; HCV\_NS2; 1.  
DR PFAM; PF01006; HCV\_NS4a; 1.  
DR PFAM; PF01001; HCV\_NS4b; 1.  
DR PFAM; PF01506; HCV\_NS5a; 1.  
DR PFAM; PF00998; HCV\_RdRP; 1.  
DR PFAM; PF01543; HCV\_capsid; 1.  
DR PFAM; PF01542; HCV\_core; 1.  
DR PFAM; PF01539; HCV\_gov; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2529 2529  
FT CARBOHYD 2788 2788  
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 13.3%; Score 80.5; DB 1; Length 3010;  
Best Local Similarity 21.1%; Pred. No. 8.4;  
Matches 26; Conservative 24; Mismatches 64; Indels 9; Gaps 3;

Qy 1 VSGSVSALRGCKKVLMEARTSTSTSDVDFVFEAVSNALLVHVRVVPVAPV-KR 59  
Db 2237 MGGNITRVESENKVVLDSDPLRAEEDEREVSAAETLRKS-----KPEPPALPIWAR 2290  
Qy 60 EQPKPAVKQDQKPK--RQASHWAVKPTAVGVHVPKPKQEALEPAQSVPOQSLEEKAL 117  
Db 2291 PDYNPPLLESKWSFDYPPVPAVHGCPPLPTTGPPIPPPKKKTIVLTSTVSSALAEATK 2350  
Qy 118 TFG 120  
Db 2351 TFG 2353  
RESULT 2  
ELAV\_DROME STANDARD; PRT; 483 AA.  
ID ELAV\_DROME

AC PL6914;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ELAV PROTEIN (EMBRYONIC LETHAL ABNORMAL VISUAL PROTEIN).  
GN ELAV.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-CANTON-S;  
RX MEDLINE-89072740; PubMed-3144044;  
RA Robinow S., Campos A.R., Yao K.-M., White K.;  
RT "The elav gene product of Drosophila, required in neurons, has three  
RL RNP consensus motifs.";  
RL Science 242:1570-1572(1988).  
RP [2]  
RP ERRATUM.  
RA Robinow S., Campos A.R., Yao K.-M., White K.;  
RL Science 243:12-12(1989).  
CC -!- FUNCTION: IS REQUIRED FOR THE PROPER DEVELOPMENT AND MAINTENANCE  
CC OF NEURONS PRESUMABLY BY AFFECTING RNA METABOLISM.  
CC -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).  
CC -!- SIMILARITY: BELONGS TO THE ELAV FAMILY OF RNP PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M21152; AAA28506.1; -.  
DR PIR; A33130; A33130.  
DR HSP; P19339; 2SXL.  
DR FLYBASE; FBgn000570; elav.  
DR INTERPRO; IPR000504; -.  
DR INTERPRO; IPR002343; -.  
DR PFAM; PF00076; Rrm; 3.  
DR PRINTS; PR00961; HODSXLRNA.  
DR PROSITE; PS01012; RRM; 3.  
DR PROSITE; PS00030; RRM\_RNP\_1; 2.  
KW RNA-binding; Repeat.  
FT DOMAIN 24 126 ALA/GLN-RICH.  
FT DOMAIN 203 210 RNA-BINDING (RNP1) (BY SIMILARITY).  
FT DOMAIN 290 297 RNA-BINDING (RNP1) (BY SIMILARITY).  
FT DOMAIN 443 450 RNA-BINDING (RNP1) (BY SIMILARITY).  
SQ SEQUENCE 483 AA; 50817 MW; E57336106310810B CRC64;  
Query Match 12.9%; Score 78; DB 1; Length 483;  
Best Local Similarity 25.9%; Pred. No. 1.8;  
Matches 28; Conservative 22; Mismatches 48; Indels 10; Gaps 2;  
Qy 10 GDGKKVLMERTSTSTSDVDFVFEAVSNALLVHVRVVPVAPVAPVAPVAPVQD 69  
Db 11 GGGVDTOAQLMQSAAAAAATAATNAAAAAPQNA-----AVAAAAQLQQQQVQAILQV 64  
Qy 70 EQPKPAVKQDQKPK--RQASHWAVKPTAVGVHVPKPKQEALEPAQSVPOQSLEEKAL 117  
Db 65 QQQQTOQ-----AVAAAAAATVQQLQQQQQAVVAAQAVVQQQQQAAAV 108  
RESULT 3  
FMR2\_HUMAN  
ID FMR2\_HUMAN STANDARD; PRT; 1311 AA.  
AC P51816; Q13521; Q14323; P78407;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)









```

DR INTERPRO: IPR002522;
DR INTERPRO: IPR002531;
DR INTERPRO: IPR002868;
DR PFAM: PF01560; HCV_NS1; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF00998; HCV_NS5b; 1.
DR PFAM: PF01543; HCV_NS6; 1.
DR PFAM: PF01542; HCV_NS7; 1.
DR PFAM: PF01539; HCV_NS8; 1.
KW Polyprotein; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydroxylase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 746
FT CHAIN 747 809
FT CHAIN 810 1026
FT CHAIN 1027 1657
FT CHAIN 1658 1711
FT CHAIN 1712 1972
FT CHAIN 1973 2420
FT CHAIN 2421 3011
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 11.8%; Score 71.5; DB 1; Length 3011;
Best Local Similarity 22.0%; Pred. No. 66;
Matches 27; Conservative 23; Mismatches 64; Indels 9; Gaps 3;

Qy 1 VSGSVSLRGDGKVKVLMKARTSTSTSDVDFVVEAVSNALLVYHVRVVPYAPV-KR 59
Db 2237 MCGNITRVESENKVVILDSFDPLVAEEDEREVSVAETILRKS-----RRFAPALPVNAR 2290

Qy 60 EQPKAVKQDEQPKRQ--ASHWAVKPTAVGVHVPVLPKKQEALEPAQSVQPSLEEKAL 117
Db 2291 PDYNPLLVETWKKPDYEPVHVHGCPLPPRSPVPPPPKRTVVLTESTLPTALAEATK 2350

Qy 118 TFG 120
Db 2351 SFG 2353

RESULT 10
H1E_STRPU
ID H1E_STRPU STANDARD; PRT; 205 AA.
AC P13375;

```

```

DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1, EARLY EMBRYONIC.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82265580; PubMed=7107576;
RA Levy S., Sures I., Kedes L.;
RT "The nucleotide and amino acid coding sequence of a gene for H1
RT histone that interacts with euchromatin. The early embryonic H1 gene
RT of the sea urchin Strongylocentrotus purpuratus."
RL J. Biol. Chem. 257:9438-9443(1982).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: EARLY EMBRYONIC.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01354; CAA24644.1;
DR HSP: P02259; IHST.
DR INTERPRO: IPR001386;
DR PFAM: PF00538; linker_histone; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 205 AA; 22173 MW; FF553FA7B9C33853 CRC64;

Query Match 11.7%; Score 71; DB 1; Length 205;
Best Local Similarity 34.3%; Pred. No. 3.2;
Matches 23; Conservative 9; Mismatches 31; Indels 4; Gaps 2;

Qy 52 VYAPVKEQPKPAVKQDEQPKRQASHWAVKPTAVGVHVP--LPKKQEALEPAQSVPOQ 109
Db 138 VKAAPKAK--KPVKKTTEKKEKKTTPKPKKPAKPAKSTPKTKPKKAAAKPKTAKPK 195
Qy 110 SLEEKAA 116
Db 196 PAXKRAA 202

RESULT 11
H1_LVTPI
ID H1_LVTPI STANDARD; PRT; 210 AA.
AC P06144;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LATE HISTONE H1.
OS Lytechinus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Temnopneustidae;
OC Lytechinus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-EMBRYO;
RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RT Lytechinus pictus and Strongylocentrotus purpuratus."
RL Nucleic Acids Res. 14:8121-8133(1986).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.

```



```
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X04488; CAA28177.1; -.
DR PIR; A25550; A25550.
DR HSP; P02259; 1HST.
DR INTERPRO; IPR001386; -.
DR PFAM; PF00538; linker_histone; 1.
DR KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
DR SQ SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;
-----
Query Match 11.7%; Score 71; DB 1; Length 210;
Best Local Similarity 31.7%; Pred. No. 3.3;
Matches 20; Conservative 11; Mismatches 20; Indels 12; Gaps 2;
Qy 56 PVKREQPKPAVKDEQKPK--KRQASHWAVKPTAVGVHVPKPKQEALEPAQSVPOQSLEE 113
| | | | | : | | | | | : | | | | : | | | : | | | :
Db 156 PAKKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAPKAAP 205
| | | | | : | | | | | : | | | | : | | | : | | | :
Qy 114 KAA 116
| | |
Db 206 KAA 208
-----
RESULT 12
TONB_HELPY
ID TONB_HELPY STANDARD; PRT; 285 AA.
AC O25899;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TONB PROTEIN.
GN TONB OR HP1341.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE-97394467; PubMed-9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey A.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venler J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
CC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE00635; AAD08383.1; -.
DR TIGR; HP1341; -.
DR KW Transport; Protein transport; Inner membrane; Periplasmic;
DR Transmembrane; Repeat.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 ANCHOR (POTENTIAL).
FT DOMAIN 36 285 PERIPLASMIC (POTENTIAL).
FT SEQUENCE 285 AA; 31585 MW; 3897A8BB7B70BDF8 CRC64;
-----
Query Match 11.7%; Score 71; DB 1; Length 285;
Best Local Similarity 25.5%; Pred. No. 4.8;
Matches 28; Conservative 18; Mismatches 48; Indels 16; Gaps 4;
Qy 8 LRGGGKVKLMEARSTSTSDVDFVFEAVSNALLVVHVRVVPVAPVREQPKPAVK 67
| | | | | : | | | | | : | | | | : | | | : | | | :
Db 35 LREDAPELAQAGT-TKVTMSLASINTNSNTKTA-----ESAKPKPEPKPKPK 82
| | | | | : | | | | | : | | | | : | | | : | | | :
Qy 68 QDEQK---PKROASHWAVKPTAVGVHVPKPKQEALEP-AOSVPOQSLEE 113
| | | | | : | | | | | : | | | | : | | | : | | | :
Db 83 KEPPKPEPKPEPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKVEE 132
| | | | | : | | | | | : | | | | : | | | : | | | :
-----
RESULT 13
H1B_CHITE
ID H1B_CHITE STANDARD; PRT; 232 AA.
AC P40276;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1B.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
RN [1]
RP SEQUENCE FROM N.A.
RA Schulze E., Wisniewski J.R., Nagel S., Gavenis K., Grossbach U.;
RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L29108; AAA21716.1; -.
DR HSP; P08287; 1GHC.
DR INTERPRO; IPR001386; -.
DR PFAM; PF00538; linker_histone; 1.
DR KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
DR SQ SEQUENCE 232 AA; 24388 MW; 60E7B69694706D00 CRC64;
-----
Query Match 11.6%; Score 70; DB 1; Length 232;
Best Local Similarity 38.8%; Pred. No. 4.7;
Matches 19; Conservative 5; Mismatches 19; Indels 6; Gaps 1;
Qy 55 APVKEQPKPAVKDEQKPKRQASHWAVKPTAVGVHVPKPKQEALEPA 103
```



Db 186 APKPAAPKPAKAKTETKPKRAAPKAKKPAAE-----KKPKAAKKPA 228

RESULT 14  
ID MYPH\_CHICK STANDARD; PRT; 537 AA.  
AC Q05623;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MYOSIN-BINDING PROTEIN H (MYBP-H) (H-PROTEIN) (86 KDA PROTEIN).  
GN MYBP.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.  
RC TISSUE-PECTORALIS MUSCLE;  
RX MEDLINE-93155224; Pubmed-7679114;  
RA Vaughan K.F., Weber F.E., Einheber S., Fischman D.A.;  
RT "Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa protein) reveals extensive homology with MyBP-C (C-protein) with conserved immunoglobulin C2 and fibronectin type III motifs.";  
RL J. Biol. Chem. 268:3670-3676(1993).  
CC -!- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH THICK MYOFILAMENTS IN THE A-BAND.  
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN THE SLOW TONIC ALD MUSCLE. NOT DETECTED IN GIZZARD OR HEART.  
CC -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL: L05605; AAA21418.1; -  
CC INTERPRO: IPR001777; -  
CC INTERPRO: IPR002965; -  
CC INTERPRO: IPR003006; -  
CC PFAM: PF00041; fn3; 2.  
CC PFAM: PF00047; ig; 2.  
CC PRINTS: PR00014; ENTYPETII.  
CC PRINTS: PR01217; PRICHEXTENS.  
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament; Repeat.  
FT DOMAIN 135 221 FIBRONECTIN TYPE-III.  
FT DOMAIN 253 312 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 331 416 FIBRONECTIN TYPE-III.  
FT DOMAIN 458 518 IG-LIKE C2-TYPE DOMAIN.  
FT CONFLICT 2 2 T -> G (IN AA SEQUENCE).  
FT CONFLICT 9 9 A -> P (IN AA SEQUENCE).  
FT CONFLICT 15 15 A -> K (IN AA SEQUENCE).  
SQ SEQUENCE 537 AA; 58678 MW; 06C4CF0EFE1DD233 CRC64;

Query Match 11.6%; Score 70; DB 1; Length 537;  
Best Local Similarity 32.3%; Pred. NO. 13;  
Matches 20; Conservative 8; Mismatches 28; Indels 6; Gaps 2;  
QY 55 APVKREOPKPAVKQ-DEQPKPKQASHWAVKPTAVGVHVPKPKQEALEPAQSVPOOSLEE 113  
Db 13 APAAKAPAPAKPAPEAPKPKPA-----PTPKEGHAPTTPKEEHAPPPKPEE 67  
QY 114 KA 115  
Db 68 HA 69

RESULT 15  
ID KPBI\_RAT STANDARD; PRT; 1242 AA.  
AC Q64649; Q64650; Q64651; Q90277;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PHOSPHORYLASE B KINASE ALPHA REGULATORY CHAIN, SKELETAL MUSCLE  
DE ISOFORM (PHOSPHORYLASE KINASE ALPHA M SUBUNIT).  
GN PKA1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RA Brushia R.J., Hove E.R., Walsh D.A.;  
RT "Fundamental physicochemical characterization of the phosphorylase b kinase holoenzyme and alpha-gamma-delta and gamma-delta subunit subcomplexes reconstituted in baculovirus-infected insect cells.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 244-267; 547-658 AND 908-1136 FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE-92381046; Pubmed-1512265;  
RA Cawley K.C., Akita C.G., Wineinger M.A., Carlsen R.C., Gorin F.A., Walsh D.A.;  
RT "Coordinated expression of phosphorylase kinase subunits in regenerating skeletal muscle.";  
RL J. Biol. Chem. 267:17287-17295(1992).  
CC -!- FUNCTION: PHOSPHORYLASE B KINASE CATALYZES THE PHOSPHORYLATION OF SERINE IN CERTAIN SUBSTRATES, INCLUDING TROPONIN I. THE ALPHA CHAIN MAY BIND CALMODULIN.  
CC -!- ENZYME REGULATION: BY PHOSPHORYLATION OF VARIOUS SERINE RESIDUES AND BY CALCIUM.  
CC -!- PATHWAY: GLYCOGEN METABOLISM.  
CC -!- SUBUNIT: POLYMER OF 16 CHAINS, FOUR EACH OF ALPHA, BETA, GAMMA, AND DELTA. ALPHA AND DELTA ARE REGULATORY CHAINS, GAMMA IS THE CATALYTIC CHAIN, AND DELTA IS CALMODULIN.  
CC -!- SIMILARITY: BELONGS TO THE PHOSPHORYLASE B KINASE REGULATORY CHAINS FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL: AF197561; AAF06673.1; -  
CC EMBL: M92917; AAA41858.1; -  
CC EMBL: M92918; AAA41859.1; -  
CC EMBL: M92919; AAA41860.1; -  
KW Glycogen metabolism; Muscle; Phosphorylation; Calmodulin-binding; Multigene family.  
FT DOMAIN 812 842 CALMODULIN-BINDING (POTENTIAL).  
FT DOMAIN 1065 1105 CALMODULIN-BINDING (POTENTIAL).  
FT MOD\_RES 974 974 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 987 987 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1009 1009 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1020 1020 PHOSPHORYLATION (BY CAPK) (BY SIMILARITY).  
FT MOD\_RES 1022 1022 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 1025 1025 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 245 245 T -> S (IN REF. 2).  
FT CONFLICT 264 265 IS -> V (IN REF. 2).  
FT CONFLICT 926 926 I -> L (IN REF. 2).  
FT CONFLICT 957 957 R -> K (IN REF. 2).  
SQ SEQUENCE 1242 AA; 139154 MW; 15E61F40D4D2E82C CRC64;

Query Match 11.6%; Score 70; DB 1; Length 1242;  
 Best Local Similarity 25.2%; Pred. No. 33;  
 Matches 30; Conservative 12; Mismatches 45; Indels 32; Gaps 4;  
 Qy 21 TSTATSVDSDVVFVAVSNALLV-----HYHVVVYAPVKREQP-----KPA 65  
 Db 680 TPTSRKGGGLDRFAAVQTTCDLMSLVAKAKELHIQNVHMYLPTKLFQFSPSLNLLDSPE 739  
 Qy 66 VKQDFQKPKRCQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPOQ-----SLEEKAAALTF 119  
 Db 740 SPQDSQVP-----SVRVEVHLPRDQSGEVDQSLVSQLKETSSIQEQADILY 786

Search completed: January 31, 2001, 15:33:10  
 Job time: 325 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:55 ; Search time 102.04 Seconds  
(without alignments)  
137.838 Million cell updates/sec

Title: US-09-301-906-5  
Perfect score: 605  
Sequence: 1 VSGSVSALRGDKKVLMEAR.....EPAQSVPOQSLEEKAAITFG 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL15.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mbc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	88	14.5	462	10	Q92RQ1	Q92rq1 arabidopsis
2	83.5	13.8	282	2	Q9RW65	Q9rw65 deinococcus
3	82.5	13.6	1820	4	Q9P2P6	Q9p2p6 homo sapien
4	80.5	13.3	453	12	Q9YK14	Q9yk14 hepatitis c
5	80.5	13.3	453	12	Q9YK13	Q9yk13 hepatitis c
6	80.5	13.3	2284	12	Q81817	Q81817 hepatitis c
7	80.5	13.3	3010	12	Q81541	Q81541 hepatitis c
8	80.5	13.3	3010	12	P89966	P89966 hepatitis c
9	79	13.1	1545	5	Q9XX31	Q9xx31 caenorhabdi
10	78.5	13.0	447	12	O89168	O89168 hepatitis c
11	78	12.9	479	5	Q9V3F6	Q9v3f6 drosophila
12	77.5	12.8	302	5	Q9VQH5	Q9vqh5 drosophila
13	77	12.7	942	13	O57452	O57452 fundulus he
14	77	12.7	1545	5	Q9VYNI	Q9vyni drosophila
15	76.5	12.6	385	2	O07110	O07110 enterococu
16	76.5	12.6	430	11	O60684	O60684 mus musculu
17	76.5	12.6	860	5	P90906	P90906 caenorhabdi
18	76	12.6	16215	5	Q9NFS3	Q9nfs3 drosophila
19	75.5	12.5	318	10	Q9SL96	Q9sl96 arabidopsis

20	75.5	12.5	451	2	Q9JWQ1	Q9jwq1 bacillus ha
21	75	12.4	1852	5	Q9VH42	Q9vh42 drosophila
22	75	12.4	3008	12	Q39929	Q39929 hepatitis c
23	74.5	12.3	2359	5	O16024	O16024 drosophila
24	74.5	12.3	3614	5	Q9W4E2	Q9w4e2 drosophila
25	74	12.2	467	13	O42465	O42465 salmo salar
26	74	12.2	1272	4	O43786	O43786 homo sapien
27	74	12.2	1311	4	O60215	O60215 homo sapien
28	74	12.2	1557	5	Q9W513	Q9w513 drosophila
29	74	12.2	1718	5	Q9NE62	Q9ne62 leishmania
30	73.5	12.1	209	13	Q02612	Q02612 gallus gall
31	73.5	12.1	933	4	Q9UQ14	Q9uq14 homo sapien
32	73.5	12.1	1026	4	Q9UJS7	Q9uj57 homo sapien
33	73.5	12.1	1179	11	Q9JK71	Q9jkr71 ratus norv
34	73.5	12.1	1272	11	O55112	O55112 mus musculu
35	73.5	12.1	3010	12	Q9J3G4	Q9j3g4 hepatitis c
36	73	12.1	713	10	Q9SK77	Q9sk77 arabidopsis
37	73	12.1	1110	13	Q91255	Q91255 petromyzon
38	72.5	12.0	438	2	O50175	O50175 mycobacteri
39	72.5	12.0	952	2	Q9RN18	Q9rn18 bacillus an
40	72.5	12.0	3011	12	O36609	O36609 hepatitis c
41	72	11.9	630	12	O66652	O66652 equine herp
42	71.5	11.8	417	3	Q9P513	Q9p513 neurospora
43	71.5	11.8	511	10	Q9SZ31	Q9sz31 arabidopsis
44	71.5	11.8	906	11	O55129	O55129 mus musculu
45	71.5	11.8	956	10	Q9LJ64	Q9lj64 arabidopsis

ALIGNMENTS

RESULT 1

Q92RQ1 ID Q92RQ1 PRELIMINARY: PRT: 462 AA.  
AC Q92RQ1: DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE 2-OXOGLUTARATE DEHYDROGENASE E2 SUBUNIT (EC 2.3.1.61).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. C24;  
RA Machuy N., Klein M., Mueller-Roeber B.;  
RT "Cloning and characterization of 2-Oxoglutarate dehydrogenase from  
RT Arabidopsis thaliana.";  
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AJ223803; CAAL1553.1;  
DR HSSP; P07016; 1E20.  
DR INTERPRO; IPR000089;  
DR INTERPRO; IPR001078;  
DR INTERPRO; IPR003016;  
DR PFAM; PF00198; 2-oxoacid\_dh; 1.  
DR PFAM; PF00364; biotin\_lipoyl; 1.  
DR PROSITE; PS00189; LIPOYL; 1.  
DR PRODOM; PD001115; 1.  
KW Transferase; Acyltransferase; Lipoyl.  
SQ SEQUENCE 462 AA; 49939 MW; 5A811153FB34EEAA CRC64;

Query Match 14.5%; Score 88; DB 10; Length 462;  
Best Local Similarity 32.1%; Pred. No. 0.32; Indels 32; Gaps 6;  
Matches 35; Conservative 12; Mismatches 30;  
QY 12 GKKVLMEARTSTSAISDVSDVDFVFEAVSNALLVHVRVYPVKRQPKPAVKQD-- 69  
Db 161 GNKV---ARISTSA-----DAVSHV-----APSEKAPERKAPKPSPP 194  
QY 70 EOKPKRQASHWAVKPTAVGVHVPLPKKQKQALEPAQSVPOQSLEEKAAALT 118









Qy 70 EQKPKRQASHWAVKPTAVGVHPLPKKQEALEPAQSVPOQSLEKAAL 117  
Db 61 QOOQTOQ-----AVAAAAAANTQLOQQQQQAVVAAQAVVQOQQQQQAAV 104

RESULT 12  
Q9VQH5 PRELIMINARY; PRT; 302 AA.  
AC Q9VQH5;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE CG2973 PROTEIN.  
GN CG2973.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID-7227;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-BERKELEY;  
RX MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J., H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.F., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Buttle H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003561; AAF51198.1; -;  
DR FLYBASE; FBgn0031467; CG2973.  
DR INTERPRO; IPR000618; -;  
DR PFAM; PF00379; insectCuticle; 1.  
DR PRINTS; PR00947; CUTICLE.  
DR PROSITE; PS00233; CUTICLE; 1.  
SQ SEQUENCE 302 AA; 32060 MW; 5DAB89E0B5D06168 CRC64;

Matches 28; Conservative 4; Mismatches 28; Indels 11; Gaps 3;  
Qy 1 VSGSVSALRGDGKKVLMPEARSTSTSDVDFVFEAVSNALLVHVHVVPYAPVKRE 60  
Db 183 VRGEYSLLDPDGYK-----RTVTYTADDVHGFGNAYNRPVALKAV----VVPVQV--A 231

Qy 61 QPKPAVKODEQ 71  
Db 232 QTPPFVARDER 242

RESULT 13  
O57452 PRELIMINARY; PRT; 942 AA.  
ID O57452  
AC O57452;  
DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE ARYL HYDROCARBON RECEPTOR 1 ALPHA.  
GN AHRI ALPHA.  
OS Fundulus heteroclitus (Killifish) (Mummichog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Cyprinodontiformes; Fundulidae; Fundulus.  
OX NCBI\_TaxID-8078;  
[1]  
RN SEQUENCE OF 117-336 FROM N.A.  
RP MEDLINE-98054305; PubMed-9391097;  
RA Hahn M.E., Karchner S.I., Shapiro M.A., Perera S.A.;  
RT "Molecular evolution of two vertebrate aryl hydrocarbon (dioxin)  
receptors (AHR1 and AHR2) and the PAS family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:13743-13748(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RA Karchner S.I., Powell W.H., Hahn M.E.;  
RT "Identification and functional characterization of two highly  
divergent aryl hydrocarbon receptors (AHR1 and AHR2) in the teleost  
Fundulus heteroclitus: Evidence for a novel subfamily of ligand-  
binding basic helix loop helix-per-arnt-sim (bHLH-PAS) factors.";  
RL J. Biol. Chem. 274:33814-33824(1999).  
[3]  
RN SEQUENCE FROM N.A.  
RA Hahn M.E., Karchner S.I., Shapiro M.A., Perera S.A.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A.  
RA Hahn M.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF024591; AAC60334.2; -;  
DR INTERPRO; IPR000014; -;  
DR INTERPRO; IPR001092; -;  
DR INTERPRO; IPR001610; -;  
DR INTERPRO; IPR003015; -;  
DR PFAM; PF00785; PAS; 1.  
DR PFAM; PF00989; PAC; 2.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
KW Receptor.  
SQ SEQUENCE 942 AA; 103682 MW; E8D041E4E61BB790 CRC64;

Query Match 12.7%; Score 77; DB 13; Length 942;  
Best Local Similarity 22.0%; Pred. No. 10;  
Matches 26; Conservative 25; Mismatches 47; Indels 20; Gaps 4;  
Qy 2 SGSVSA-LRGDGKKVLMPEARSTSTSDVDFVFEAVSNALLVW-----HYHRV 52  
Db 81 SNGAAAEARGNGDE-----TTGAAEKQISEGELLQLQALNGFVLVITADGTIFYSSHTIQ 134

Qy 53 PY-----APVKREQPKPAVKODEQPKRQASHWAVKPTAVGVHPLPKKQEALEPAQS 105  
Db 135 DYLGFGHOTDVHQSAVELVHTEDQQLARNLHNLWALNPPQAAASVISQSSQEMEPDSS 192



```
RESULT 14
Q9VYN1 PRELIMINARY: PRT; 1545 AA.
AC Q9VYN1:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CG11245 PROTEIN.
GN CG11245.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.W., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003489; AAF48161.1; -.
DR FLYBASE; FBgn0030388; CG11245.
SQ SEQUENCE 1545 AA; 166082 MW; 2ABCEABED136C8B8 CRC64;
```

```
Query Match 12.7%; Score 77; DB 5; Length 1545;
Best Local Similarity 24.5%; Pred. No. 19;
Matches 36; Conservative 16; Mismatches 31; Indels 64; Gaps 6;

QY 19 ARTSTSATSDVDEVDVFEAVSNALLVHVHVVPVAPVKRQPKPAVKDEQKPKRQAS 78
DB 138 ARPAINATADID-----VSSS-----RYRAVPSKSOKEQPKPEVKQKQAE 185
QY 79 HWAVKP-----TAVGVHVPL-----PKQEE-- 98
DB 186 ---VKPSKETPCSSDASAEIRPNRSTIRNRPNRVRRLSTIRRSKDRSAESPKEQVQ 242
```

```
QY 99 -----ALEPAQSVPPQOSLEEKAA 116
DB 243 EHOFRLPPATEPSLSWROKLAELAA 269

RESULT 15
O07110 PRELIMINARY: PRT; 385 AA.
AC O07110:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CELL DIVISION PROTEIN.
GN DIVIB.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A24836;
RX MEDLINE=97431524; PubMed=9287029;
RA Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E.,
RA Dougherty T.J.;
RT "Identification and characterization of cell wall-cell division gene
RT clusters in pathogenic gram-positive cocci.";
RL J. Bacteriol. 179:5632-5635(1997).
DR EMBL; U94707; AAC45637.1; -.
KW Cell division.
SQ SEQUENCE 385 AA; 43741 MW; 1BFB17496D980D5A CRC64;

Query Match 12.6%; Score 76.5; DB 2; Length 385;
Best Local Similarity 31.3%; Pred. No. 4.2;
Matches 21; Conservative 15; Mismatches 26; Indels 5; Gaps 2;

QY 53 PTAPVKRQPKPAV----TQDEQKPKRQASHWAVKVTAVGVHVPLP-KKQEALEPAQSV 107
DB 16 PLPPKKSESOPELTPWQKNQYLLKKQAEAAKGENEAQEVTTILQEQSQEPEQQLP 75
QY 108 QOSLEEK 114
DB 76 QETVEEE 82

Search completed: January 31, 2001, 15:32:37
Job time: 402 sec
```

**This Page Blank (uspto)**

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      January 31, 2001, 15:28:18 ; Search time 42.21 Seconds
              (without alignments)
              1812.166 Million cell updates/sec

Title:       US-09-301-906-15
Perfect score: 11520
Sequence:    1 MDYTRPLRVFSFPHVNTLE.....QSVSDALLHTFAPACCFRGI 2237

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched:   268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      7268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database :

A\_Geneseq\_36.\*

1:	/SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	11520	100.0	2237	21	Y58148	GLRaV-3 polyprotein	
2	5247	45.5	1390	18	W21636	Grapevine leafroll	
3	1495	13.0	291	21	Y58145	GLRaV-3 polyprotein	
4	1432	12.4	272	21	Y58144	GLRaV-3 polyprotein	
5	794.5	6.9	2639	20	W73476	Grapevine leafroll	
6	631.5	5.5	309	21	Y91969	PMWaV-2 helicase.	
7	605	5.3	120	21	Y58143	GLRaV-3 polyprotein	
8	323	2.8	598	21	Y91963	PMWaV-1 helicase.	
9	323	2.8	1128	11	R05107	Sequence encoded b	
10	309	2.7	1116	15	R60608	Tobamovirus replic	
11	290	2.5	2161	20	W87724	Rupestris stem pit	
12	274.5	2.4	992	11	R06531	Protein encoded by	

13	233	2.0	1704	15	R49657	Sequence of Helioth
14	185	1.6	2039	19	W53222	Haemophilus paraga
15	184.5	1.6	1693	19	W70368	Hepatitis E virus
16	184	1.6	1698	20	Y31381	HEV-US1 ORF1 prote
17	179.5	1.6	1693	15	R51264	HEV strain protein
18	179.5	1.6	1693	17	R81813	Hepatitis E virus
19	179.5	1.6	1693	19	R91519	Hepatitis E virus
20	176	1.5	1708	20	Y31384	HEV-US2 ORF1 prote
21	174	1.5	1708	20	W93408	Swine HEV ORF 1 pr
22	169.5	1.5	1693	19	W80196	Protein encoded by
23	169.5	1.5	1693	19	W71209	Protein encoded by
24	160.5	1.4	1693	12	R14618	Protein encoded by
25	158	1.4	1946	19	W47273	Lactobacillus bulg
26	154.5	1.3	1532	10	P93357	Sequence of the ca
27	152.5	1.3	1705	10	P94365	Sequence of part o
28	151.5	1.3	1706	11	R08031	Adenyl cyclase fro
29	146.5	1.3	1194	20	Y37825	Chlamydia trachoma
30	143.5	1.2	1252	16	R80530	B. sphaericus SLP.
31	142.5	1.2	1981	16	W42634	Protein sequence t
32	141.5	1.2	2353	17	R93933	HEMOPHILUS ADHESI
33	140	1.2	800	20	Y08624	Human secreted pro
34	140	1.2	1233	19	W41935	Nudaurelia beta-li
35	140	1.2	1233	19	W34536	Nudaurelia beta vi
36	140	1.2	3647	11	R05041	Filamentous haemag
37	138.5	1.2	2032	20	Y00238	Enterococcus faeca
38	138.5	1.2	2032	20	Y00240	Enterococcus faeca
39	138.5	1.2	2032	20	Y00242	Enterococcus faeca
40	137	1.2	928	20	W88417	Chlamydia pneumoni
41	137	1.2	949	20	Y35060	Chlamydia pneumoni
42	137	1.2	3596	21	Y87407	Bordetella pertuss
43	136.5	1.2	1978	20	Y27230	Amino acid sequenc
44	136.5	1.2	2285	20	R98149	Bacillus subtilis
45	135.5	1.2	1302	16	R84741	Multidrug resistan

## ALIGNMENTS

RESULT	1	
Y58148		
ID	Y58148	standard; Protein; 2237 AA.
XX		
AC	Y58148;	
XX		
DT	07-MAR-2000	(first entry)
XX		
DE	GLRAV-3	polyprotein.
XX		
KW	Grapevine; leafroll;	grapevine leafroll-associated virus 3; GLRAV-3;
KW	viral disease; yield loss;	sugar content; inhibition; infection;
KW	replication; polyprotein;	domain; proteinase; methyltransferase;
KW	helicase; RNA-dependent;	RNA polymerase; untranslated region;
KW	transgenic plant;	component; resistant; truncation; deletion; antisense;
KW	expression; detection;	antibody.
XX		
OS	Grapevine leafroll-associated virus 3.	
XX		
Key	Location/Qualifiers	
FH	85..204	
Domain	/note= "Proteinase domain (Y58143)"	
FT	460..731	
FT	/note= "Methyltransferase domain (Y58144)"	
FT	1922..2212	
FT	/note= "Helicase domain (Y58145);"	
FT		
XX		
XX	WO9955880-A1.	
PN		
XX		
PD	04-NOV-1999.	
XX		
XX	29-APR-1999;	99WO-US09307.
PF		
XX		
XX	29-APR-1998;	98US-0083404.
PR		
XX		

PA (CORR ) CORNELL RES FOUND INC.  
XX  
PI Consalves D, Ling K;  
XX  
XX WPI; 2000-062035/05.  
DR N-PSDB: 249208.  
DR  
XX Newly isolated grapevine leafroll virus protein or polypeptide useful  
PT for producing transgenic plants conferring viral disease resistance -  
PT  
XX  
XX Claim 3; Fig 16; 84pp; English.  
XX  
CC This sequence represents the polyprotein encoded by open reading  
CC frame (ORF) 1a from the grapevine leafroll-associated virus 3  
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever  
CC grapes are grown. Although not lethal, it causes yield losses and  
CC reduction in sugar content. The virus encodes several proteins,  
CC which may serve as targets for the inhibition of viral infection or  
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),  
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase  
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase  
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF  
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 1i.  
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'  
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are  
CC useful for the generation of transgenic plants and plant components.  
CC Such transgenic plants may be resistant to viral disease, for example,  
CC this property being conferred on the plants via the use of nucleotides  
CC encoding truncated or internally deleted proteins, or via the use of  
CC antisense nucleotides to inhibit viral gene expression. The nucleotides  
CC may additionally be used in the detection of viral nucleic acids in a  
CC tissue sample. The proteins can be used to generate antibodies which  
CC may be used to detect GLRaV-3 in plant samples. The isolation of  
CC GLRaV-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.  
XX  
XX Sequence 2237 AA:  
SQ

Query Match 100.0%; Score 11520; DB 21; Length 2237;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDYIRLRVFSFPHVNTLEYRYNKANGDVGAFLLTMKFIGNVKLSDFTPRCAAMIYIG 60  
DB 1 mdyirrvfsfphvntleyrynkangdvgaflltmkfignvklsdftprcaamiyig 60  
QY 61 KLTKGVKRTFVPPVKCFARQYAVVSSVSALRGDGGKVLMEARTSATSSDVDFDVF 120  
DB 61 kltkgvkrtrfvppvkcfarqyavvssvsalrgdggkvlmeartstsatsdvsdfdvf 120  
QY 121 EAVSNALLVVHYHRVVPYAPVKRBPQPAVKQDBQPKROASHWAVKPTAVGVHVPKPK 180  
DB 121 eavsnallvvhyhrvvpypavkrebqpvavkqdeqbkrgashwavkptavgvhvpkpk 180  
QY 181 QEALPEAQSPQOSLEEKAAALTFGLFFSKGGDESDAVILRKGKLFNRALNPIDVKNTE 240  
DB 181 qealepaqspqosleekaaaltfglffskgggdesdaviilrkgklnrlnvpidvknntf 240  
QY 241 VWAKINDEASRRRGYFVVKORAVKFFPIVGRATIEDFIYNTAPGCDVALPRIELWSMRE 300  
DB 241 vwakiwdeasrrrgyfvyvkdravkffpivgratiedfiyntapgcdvalprielwsmre 300  
QY 301 RAFVCTTKGWCWFNNRLRGIEIVRRRCFSSFSIGFLMHLCFRSLKVIRFAGTNILHMPS 360  
DB 301 rafvcttkgwcwfnnrlrgieivrrrcfssfsigflmhlgfrslkvirfagtnilhmps 360  
QY 361 LNEERTFGWKGGDYLLPNVPKTAIVAGDRTLGLGEILLASVANALNQBEVYSSVSSITNR 420  
DB 361 lneertfgwkggdyllpnvpktaivagdrtlglgeilllasvanalnqbevyssvssitnr 420  
QY 421 LVLRDQSALLSHLDTKLCDFSQSDAMIREKPSHRCDDVFLKPREREKLRLEFPFELSIOFS 480  
DB 421 lvlrdsqallshldtklcfdsqsdamirekpsshrccdvflkprereklrlefpelsiofs 480

DB 421 lvlrdsqallshldtklcfdsqsdamirekpsshrccdvflkprereklrlefpelsiofs 480  
QY 481 DSVRSHHPFANAMRSCFNIGFSRRCGNVCFPDIGGSFTYHVKAGHVNCNPVLVDVKDV 540  
DB 481 dsvrshhpfanamrscfnigfsrrcgvncvfdiggsftyhvkgahvncnvpvlvdvkdv 540  
QY 541 KRRINEILFLSTAGSDSVSSDLLETEAASKSVSYCSRESQNCDSRADAGFMDVDYDISPQ 600  
DB 541 krrineilflstaggdsyvssdlleteaasksvsycsresqncdsradagfmdvdydispq 600  
QY 601 QVAEAMDKKALVFDIALMFPVELLYNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEYH 660  
DB 601 qvaeamdkkalvfdialmfpvellyngevyleeldtlvkrergyaynvvgcgemeyeh 660  
QY 720 SFSNVSGFFPYFYSVRTSSGNVFKLEYEGYRGYHHLNMCRAQKSPGTEVTYVRSVLVPSVVG 720  
DB 720 sfsnvsgffpyfysvrtssgnvfkleyegyrcgyhhlmcraqkspgtevtyvrsvlvpsvvg 720  
QY 780 KSLVFIPIVWAGSSVSFKTIVLDSDFVRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVI 780  
DB 780 kslvfiipvwagssvsfktivldsdfvriysyalntigtfenrtfeyavgavrsqkthvi 780  
QY 840 TGSRVVHSKVDISPDDMWGLVAVMAQAIDRAKSIRSYNFIKASEGSLAGVKFLFFQTV 840  
DB 840 tgsrvvhskvdispdmdwglvavmaqaikdraksirsynfikasegslagvklffqtv 840  
QY 900 GDCFSNAVSVYAKAMVDNFNVLTLMSMPRAFTRKVPVGVVVTICTSGASDRLELRGAF 900  
DB 900 gdcfsnavsvyakamvdnfnvltlmsmpraftrkvpvgvvvtictsgasdrlelrgef 900  
QY 960 DISKETGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNV 960  
DB 960 disketgrklknslrvfsvraivedsikvmkamktedgkplpitedsvyafimgnvsnv 960  
QY 1020 HCTRAGILGGSKATVSVSSKGLVARGAAYKAFSGITSFFSTGSLFYDRGLTEDELDAL 1020  
DB 1020 hctragilggskatvsvsskglvargaatkafsgitsffstgslfydrglteiderdal 1020  
QY 1080 VRTENAINSPVGILETSRVAVSKVVAGTKFEPWSEVSLNDFTFVLRNKLVLIGIVASLGA 1080  
DB 1080 vrtenainspvgiletssrvavskvvagtkfepwsevslnndftfvlrnlkvliligivaslga 1080  
QY 1140 APIAWKYRRGIAANARYAGSSYETLSLSQAAGGLRGLTSSVSGSLVYVRGFSASV 1140  
DB 1140 apiawkyrriigaanaryagssyetlsslssaagglrgltsstsvsgslvyrvgfssav 1140  
QY 1200 TVTRATVAKRQVPLALLSFTSYAISGCSMLGIWAHALPRHLMFFGLGLTLLGARASANT 1200  
DB 1200 tvtratvakrqvplallsfstysaigscsmllgiwahalprhlmlffglgltilgarasant 1200  
QY 1260 WKFGCFNNMCAPVEVVMRGKSVSSLLPITLGVSLIIRGLLNDTIPQLAVYPPVEGRNV 1260  
DB 1260 wkfgcfnnmcavpevvvmrgksvssllpiltlgvslirglndtipqlayppvegrnv 1260  
QY 1320 YDETLRYRDFDYDEGAGSGTQHEAVPGDDNDGTSVSSYDVVTNVRDVGISTNGEVT 1320  
DB 1320 ydetlryrddfdydegagsgtqheavpgdndgstssvssydvvtnvrdvgistngevt 1320  
QY 1380 GEEETHSPRSVQYTYVEEVAPSAVAERQDPSGSGTADAMAFVESVKKQVDVDFHQQS 1380  
DB 1380 geeethsprsvqytyveevapsaavaerqdpssgsgtadamaftvesvkkqvdfvfhqqs 1380  
QY 1440 SGETARVEVDGKLLPESVVGAEPTQERGAADGNTAQTAVNEGDEPQVSSVLSSPOA 1440  
DB 1440 sgetarevevdgkllpesvvgaeptqergraadgntaqtaavnegdepvqssvlssppqa 1440  
QY 1500 DIPKVTQSEVHAQKOEIVPLATVSGATPIVDEKPAFVYTRGVTKIIDKGAHAHVAEK 1500  
DB 1500 dipkvtqsevhakoeivplatvsgatpivdekpaftvtrgvtkiidekgahavaek 1500  
QY 1560 KOVOVEQPKORSLTINGKACKQLCMERTCSGQVOLDVNEATTATPFSNAFTFVDNLKG 1560  
DB 1560 kvoveqpkorsltingkackqlcmertcsgqvoldivneattatpfsnaftfvdnlkg 1560

QY 1561 RSAPFSLGEGYTYNGSHVSGWPRALIEDILTALKYPSVFDHCLVQYKMGGGVPFFHA 1620  
 Db 1561 rsavfslggytyngshvsgwpralediltalkypsvfdhclvqykmggvypfha 1620  
 QY 1621 DDEECYPSDNPILTNLVGNKANSFKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNS 1680  
 Db 1621 ddeecypsdnpiiltvnlvganfctckrgkvmvinvasgdyflmpcgfqrthlshvns 1680  
 QY 1681 IDEGRISLTFRATRRVFGVGRMLQIAGVSDKSPGVPNQPOSGQATRTITPKSGKAL 1740  
 Db 1681 idegrisltrfratrrvfgvgrmlqagvsdekspgvpnpqpsqgatr itcpksgkal 1740  
 QY 1741 SEGSGREYKGRSTYSWCEQDYVRKCEWLRADNPWALEPDYTPMTFEVVKGTSEDVAV 1800  
 Db 1741 segsgrevkgrstyslwceqdyvrkcewlradnpwalepdypmtfevvtgtsedavv 1800  
 QY 1801 EYLKYLAIIGIERTYRALLMARNIATVTAEGVLKVPNOVYESLPGFHYKSGTDLIFHSQ 1860  
 Db 1801 eylkylaiigiertyrallmarniavttaegvlkvpnpqvyeslpghfhyksgtdl ifhstq 1860  
 QY 1861 DGLVRDLPVYLIAEKGIPTCKGDVADVVALGDNLFVCDIILVFHDAINLIGALKVARCG 1920  
 Db 1861 dglvrldlpvyliaekgip tckgdavadvvalgdnlfvcddilvfhdainligalkvarcg 1920  
 QY 1921 MYGESFKSEYKCYNAPGGGKTTTLVDFEYKSPNSTATITANVGSSDINNAVKKRDPN 1980  
 Db 1921 mygesfkseycynapgggktttlvdfeykspnstatitanvgssedinnavkkrdpn 1980  
 QY 1981 LEGLANSATTNRSRVNFIVRGMYKRVLVDEVMHMQGLLOLVFATGASEGLFFGDI NOI 2040  
 Db 1981 leglsattvnsrvnfi vrgmykrvlvdevmhmqgllolvfatgaseglffgdingi 2040  
 QY 2041 PFINREKVFMDCAVFPVKESVYTSKRYCPLDVCVLLSMTVRGTEKCYPEKVVSGK 2100  
 Db 2041 pfinkekvfmdcavfpvkessvytskrycpldvcvllsmtvrgtekcypekvvsqk 2100  
 QY 2101 DKPVRSLSKRPICTDDVAEINADVYLCTOLEKSDMKRSKLGKGTETPMTVTHEAQGK 2160  
 Db 2101 dkpvrslskrpigtddvaeinadvylctoleksdmkrskkgketpmtvtheagqk 2160  
 QY 2161 TFSDDVLFRTKKADDSLFRTKOPHILVGLSRHRSRLSVYALSSKLDKDKVGTIYISDASPOSV 2220  
 Db 2161 tfsdvvlftrtkadddlfrtkophilvglsrhrrslsvyaa lssklddkvgtiyisdaspsv 2220  
 QY 2221 SDALLHTFAPAGCFRGI 2237  
 Db 2221 sdallhtfapagcfrgi 2237

RESULT 2  
 W21636  
 AC W21636 standard; Protein; 1390 AA.  
 AC W21636;  
 DT 19-SEP-1997 (first entry)  
 DE Grapevine leafroll virus helicase.  
 KW GLRaV; grapevine; Vitis; rootstock; leafroll; disease resistance;  
 KW transgenic plant; tristeza virus; citrus; helicase.  
 OS Grapevine leafroll associated virus type 3 isolate NV1.  
 PN W09272700-A2.  
 XX 26-JUN-1997.  
 XX 20-DEC-1996; 96WO-US20747.  
 XX 21-DEC-1995; 95US-0009008.  
 XX

(CORR ) CORNELL RES FOUND INC.  
 Gonsalves D, Ling K;  
 WPI; 1997-341691/31.  
 N-PSDB; T72214.  
 DNA encoding grape-vine leaf-roll virus proteins - useful to impart  
 viral-resistance to Vitis scion or root-stock cultivar(s)  
 Claim 4; Page 69-75; 172pp; English.  
 A helicase (W21636) was identified as the expression product of  
 open reading frame 1a (T72215) of a double-stranded cDNA molecule  
 derived from grapevine leafroll associated virus type 3 (GLRaV-3)  
 isolate NY-1 dsRNA. The helicase, and other GLRaV-3 polyptides  
 (see also W21637-47), can be expressed in transgenic plants as a  
 means of imparting GLRaV resistance to Vitis scion or rootstock  
 cultivars or tristeza virus resistance to citrus scion or rootstock  
 cultivars, or used to raise antibodies useful for detection of  
 GLRaV.  
 Sequence 1390 AA;  
 Query Match 45.5%; Score 5247; DB 18; Length 1390;  
 Best Local Similarity 74.1%; Pred. No. 0;  
 Matches 1055; Conservative 96; Mismatches 206; Indels 66; Gaps 17;  
 QY 848 VSVYAKAMVHDNFVNLETLMSPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKET 907  
 Db 1 vsvyakamvhdnfnletlmspf irkvpgsvvvtictsgasdrlelr gafdsketf 907  
 QY 908 GRKLNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVNVHCRAGL 967  
 Db 61 srlrnsrlrvfsraivedsikvmkamedgkplpiedsvyafimgnvnhcrag l 120  
 QY 968 LGGSKATVVSVSKGLVARGAATKAFSGITSTFSTGSLFYDRGLTEDERLDALVRTENA 1027  
 Db 121 lggskacaaslavkgaasratgkifsgitstfslsagglfydeglt pgerldalttrrehav 180  
 QY 1028 NSPVGTLTSTRVAVSKVAVGTKEFWSEVSLNDFTFVLRKNKVLIGIFVASLGNAPTAWKY 1087  
 Db 181 nspvgtltrsvavskvavgtkefwsevslnndfttfvlnknkvl igifvaslgnaptawky 240  
 QY 1088 RRGIAANARRYAGSSYETLSLSQAAAGLRLTSTVSGSLVVRGRFSSAVTVTRATV 1147  
 Db 241 rriartgvdvfhrarsgtaaiglcisggrslagdaargaitvtrgglssavavtrntv 300  
 QY 1148 AKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFELGLTLLGARASANTWKEGFS 1207  
 Db 301 arrqvplallsfstsyavsgctllgiwahalprhlmffelgltlfgvsasstswslggyt 360  
 QY 1208 NNMCAVPEVWRGKSVSSLLLPITLGVSLIIRGLLNDTTPOLAYVPPVGRNVYDITLRY 1267  
 Db 361 nslftvpeltwegrysrilpqeaigslivrrgllsetvpqityvpplegrvnyddaln f 420  
 QY 1268 YRDFDYDEGAGSPGTQHEAVPGDNDGSTSVSSYDVVTNVRDVGITSTNG-----EV 1319  
 Db 421 yrdfyddgagsgptagqsdpg-tntsdtsvsvfs-----ddgip asgggfdarvesg 471  
 QY 1320 TGEETHSPR-SVQYTYVEE-EVAPSAVAERQGD---PSGSGTADA-----MAFVESG 1368  
 Db 472 pshavdesprgsvefvyrvrvdehpacgeaekdlitplgtavlesppvgpeagsapnv 531  
 QY 1369 KKGVDVDFHQSSGETAREVEVDGKGLLPESVVGEPAPTOERGRAADGNTAQTA VNEGDR 1428  
 Db 532 edgcpev-----eaekcsevlvd-----vpss---epvqvdelestng--vqaarte---e 574  
 QY 1429 PVQSSILVSSPOADIPKVTQSEVHAQKVEKQVDP-----LATVSGA-----TPIVDEK 1475  
 Db 575 vvvqgdtcga-----gvaksev-sqrvfpapqpaheagleassgavvleqlvsvpvavak 627

Qy 1476 PAPSV-TTRGVKKIIDKGAHVAHEKQVQVEQPKQKSLTINEGKAGKQKCMFRTCSGV 1534  
Db 628 tvlsvekaelkavdkgavvhakevnpvktlprgalkisedvtrkelcmftrcscgv 687  
Qy 1535 QLDVYNEATATRESNAFTFDNLKGSAYFESKLGEBYTYNGSGSHVSSGHPRALEDILT 1594  
Db 688 qldvyneatatrisnaftrfvsdkgrsavfisklgegytynggshvssgwpdraledilt 747  
Qy 1595 AIKYPVSFDHCLVOKYKMGCGVPPHADDEECYPSDNPILTIVNLVKANFSTKCRKGGKVM 1654  
Db 748 aikypvsfdhclvokymgvgppfhadeecypsdnpiiltvnlvgkanfstckrkggkvm 807  
Qy 1655 VINVASGDYFLMPCGFORTHLSVNSIDEGRISLTFRATRRVFGVGRMLQAGVSDSKS 1714  
Db 808 vinvasgdyflmpcgfqrthlsvnsidegrisitfratrrvfgvgrmlqlagvssdeks 867  
Qy 1715 PCVPNQPOQOGATRTTPKSGGKALSEGSGREYKSGSTYSIWCFODYVVRKCEWLRADNP 1774  
Db 868 pcvpnqpqsggatrtpksggkalsegsgrevkgrstysiwceqdyvrkcewlradnp 927  
Qy 1775 VMALEPDYTPMTFVWVKTGTSDEDAVVEYLYLAIGIERTYRALLMARNIAVTTAEGVLKV 1834  
Db 928 vmaikpytpmtfevkvagtsedavveylkylagigrtyrallmarniavtttaegvlkv 987  
Qy 1835 PNQVYESLPGFHVYKSGTDLIHFSTQDGLVRDLPYVLIIEKGIPTFKGKDVADVVALGDN 1894  
Db 988 pnqvyeslpgfhvyksgtdlihfstqdgllrvrdlpyvfiaekgifikgkdvadvvalgdn 1047  
Qy 1895 LFVCDLILVHDAINLCAKVARCGMVGESFESKSFYKCYNAPPGGGKTTTLVDFEVKSP 1954  
Db 1048 lsvcdlllvfhda inlmgalkvarcgmvgesfksfeykcy nappgggkttlmlvdefvksp 1107  
Qy 1955 NSTATITANVGSSBDINMAVKKRPNLGLNSATTVNSRVVNFIVRGMKRVLVDEVHMM 2014  
Db 1108 nstati tanvgssedimavkkrpnlgl nsattnsvrvnfvirgmykrvlvdevymm 1167  
Qy 2015 HQGLQLGVFATGASEGLFFGDIQIPIFNREKVFMDCAVFPKESVYTSKSRCLP 2074  
Db 1168 hqgilqlgvfatgaseglffgdiqipinrekvfmdcavfpkkesvvytsksrclp 1227  
Qy 2075 DVCYLLSSMTVRGTEKCYPEKVVSGDKPVRSLSKRPITGTTDDVAEINADVYLCMTQLE 2134  
Db 1228 dvcyllssmtvrgtekyekpekvsvsgdkpvrslskrpigttdtdvaeinadvylcmtqle 1287  
Qy 2135 KSDMKRSLKGGKETPVMTVHEAOGKTFSDVVLFRTRKADDSLTFTKPHILVGLSRHTRS 2194  
Db 1288 ksdmksrlkggketpvmvtvheaogktsdvvlfrtrkaddslftkphilvglshrtrs 1347  
Qy 2195 LVYAALSKLDDKVGTYISDASPOSVSALLHTFAPAGCFRGI 2237  
Db 1348 lvyaalselddkvgtysisdasposvsallhtfapagcfrgi 1390  
RESULT 3  
Y58145 ID Y58145 standard; Protein; 291 AA.  
XX AC Y58145;  
XX AC Y58145;  
DT 07-MAR-2000 (first entry)  
XX GLRAV-3 polyprotein helicase domain.  
XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;  
KW viral disease; yield loss; sugar content; inhibition; infection;  
KW replication; polyprotein; domain; proteinase; methyltransferase;  
KW helicase; RNA-dependent; RNA polymerase; untranslated region;  
KW transgenic plant; component; resistant; truncation; deletion; antisense;  
KW expression; detection; antibody.  
OS Grapevine leafroll-associated virus 3.  
XX  
PN W09955880-A1.

XX 04-NOV-1999.  
XX 29-APR-1999; 99WO-US09307.  
XX 29-APR-1998; 98US-0083404.  
XX (CORR ) CORNELL RES FOUND INC.  
XX Gonsalves D, Ling K;  
XX WPI; 2000-062035/05.  
DR N-PSDB; 249205.  
XX  
PT Newly isolated grapevine leafroll virus protein or polypeptide useful  
PT for producing transgenic plants conferring viral disease resistance -  
PS Claim 15; Fig 11; 84pp; English.  
XX  
CC This sequence represents the helicase domain of the  
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3  
CC (GLRAV-3). Leafroll is a serious viral disease, occurring wherever  
CC grapes are grown. Although not lethal, it causes yield losses and  
CC reduction in sugar content. The virus encodes several proteins,  
CC which may serve as targets for the inhibition of viral infection or  
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),  
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase  
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase  
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF  
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.  
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',  
CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are  
CC useful for the generation of transgenic plants and plant components.  
CC Such transgenic plants may be resistant to viral disease, for example,  
CC this property being conferred on the plants via the use of nucleotides  
CC encoding truncated or internally deleted proteins, or via the use of  
CC antisense nucleotides to inhibit viral gene expression. The nucleotides  
CC may additionally be used in the detection of viral nucleic acids in a  
CC tissue sample. The proteins can be used to generate antibodies which  
CC may be used to detect GLRAV-3 in plant samples. The isolation of  
CC GLRAV-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.  
XX  
SQ Sequence 291 AA:  
  
Query Match 13.08; Score 1495; DB 21; Length 291;  
Best Local Similarity 100.0%; Pred. No. 5.1e-113;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1922 VGESFKSFYKCYNAPPGGGKTTTLVDFEVKSPNSTATITANVGSSBDINMAVKKRPNL 1981  
Db 1 vgesfksfeykcy nappgggktttlvdfevkspnstatitanvgssedimavkkrdpnl 60  
  
Qy 1982 EGLNSATTVNSRVVNFIVRGMKRVLVDEVHMMHQGLQLGVFATGASEGLFFGDIQIPI 2041  
Db 61 eglnsattvnsrvvnfvirgmykrvlvdevhmmhqqllqlgvfatgaseglffgdiqip 120  
  
Qy 2042 FINREKVFMDCAVFPKESVYTSKSRCLPDLVCYLLSSMTVRGTEKCYPEKVVSGDK 2101  
Db 121 finrekvfmdcavfpkkesvvytsksrclpdlvcyllssmtvrgtekyekpekvsvsgdk 180  
  
Qy 2102 KPVVRSLSKRPIGTDDVAEINADVYLCMTQLEKSDMKRSLKGGKETPVMTVHEAOGK 2161  
Db 181 kpvvrslskrpigtddvaeinadvylcmtqlksdmkrsllkggketpvmvtvheaogkt 240  
  
Qy 2162 FSDVVLFRTRKADDSLTFTKPHILVGLSRHTRSILVYAALSKLDDKVGTYI 2212  
Db 241 fsdvvlfrtkaddslftkphilvglshrtrsilyaalsskllddkvgtyi 291  
  
RESULT 4  
Y58144

ID	Y58144	standard; Protein; 272 AA.
XX	XX	
AC	AC	Y58144;
XX	XX	
DT	07-MAR-2000	(first entry)
XX	XX	
DE	GLRaV-3	polyprotein methyltransferase domain.
XX	XX	
KW	Grapevine; leafroll;	grapevine leafroll-associated virus 3; GLRaV-3;
KW	viral disease; yield loss;	sugar content; inhibition; infection;
KW	replication; polyprotein;	domain; proteinase; methyltransferase;
KW	helicase; RNA-dependent;	RNA polymerase; untranslated region;
KW	transgenic plant; component;	resistant; truncation; deletion; antisense;
KW	expression; detection;	antibody.
XX	XX	
OS	Grapevine leafroll-associated virus 3.	
XX	XX	
PN	WO9955880-A1.	
XX	XX	
PD	04-NOV-1999.	
XX	XX	
PF	29-APR-1999;	99WO-US09307.
XX	XX	
PR	29-APR-1998;	98US-0083404.
XX	XX	
PA	(CORR )	CORNELL RES FOUND INC.
XX	XX	
PI	Consalves D, Ling K;	
XX	XX	
DR	WPI; 2000-062035/05.	
XX	XX	
DR	N-PSDB; 249204.	
XX	XX	
PT	Newly isolated grapevine leafroll virus protein or polypeptide useful	
XX	XX	for producing transgenic plants conferring viral disease resistance -
PS	Claim 5; Fig 8; 84pp; English.	
XX	XX	
CC	This sequence represents the methyltransferase domain of the	
CC	polyprotein (Y58148) from the grapevine leafroll-associated virus 3	
CC	(GLRaV-3). Leafroll is a serious viral disease, occurring wherever	
CC	grapes are grown. Although not lethal, it causes yield losses and	
CC	reduction in sugar content. The virus encodes several proteins,	
CC	which may serve as targets for the inhibition of viral infection or	
CC	replication. These proteins include the 242-248 kD polyprotein (Y58148),	
CC	encoded by open reading frame (ORF) 1a, and which comprises a proteinase	
CC	domain (Y58143), a methyltransferase domain (Y58144) and a helicase	
CC	domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF	
CC	1b; and a protein of unspecified function (Y58147), encoded by ORF 1l.	
CC	Nucleotides encoding these proteins, or fragments thereof, and the 5'	
CC	and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are	
CC	useful for the generation of transgenic plants and plant components.	
CC	Such transgenic plants may be resistant to viral disease, for example,	
CC	this property being conferred on the plants via the use of nucleotides	
CC	encoding truncated or internally deleted proteins, or via the use of	
CC	antisense nucleotides to inhibit viral gene expression. The nucleotides	
CC	may additionally be used in the detection of viral nucleic acids in a	
CC	tissue sample. The proteins can be used to generate antibodies which	
CC	may be used to detect GLRaV-3 in plant samples. The isolation of	
CC	GLRaV-3 facilitates the production of agents that reduce the risk of	
CC	infection or damage by the virus in vineyards.	
XX	XX	
SQ	Sequence	272 AA;
Query Match 12.4%; Score 1432; DB 21; Length 272;		
Best Local Similarity 100.0%; Pred..No. 5.9e-108;		
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps		
QY	460	LKPREKRLPELSIQFSDVSRSHPPANAMRSCNCFISRRCCNVCFDGGSFY 519
Db	1	lkprekrlrelpelsiqfsvrshppanamrscngifsrccnvcvfdggstfy 60
QY	520	HVAGHGVNCVCPVLVDVKVKRRINEILFLSTPAGDSVSSDLLTEAAKSKSVSYCSRES 579

Db	61	hvkaghvncvnpvldvdkvrrineillstaggdsyvssdlleatasksvsycsres	120
Qy	580	QNCDSRADAGFMVDVYDISPOQVAEAMDKKALVFDIALMFPVELLYGNGEVYLELDTL	639
Db	121	qncdsradagimdvdispqvaeamdkgalvfdialmfpvellygngevyleldtl	180
Qy	640	VKREGDYLAYNVGCGEMYEHSFSNNVSGFFTFYSVVRTSSGNVFKLEYRCGYHHLTMC	699
Db	181	vkregdylaynvvgcgemyehsfnsnvgfftfysvrtssgnvfkleyrcgyhhltnmc	240
Qy	700	RAQSPGTEVYRSLVPSFVGKSLVFPVAVG	731
Db	241	raqspgtevtyrslvpsfvgkslvfpvavg	272
RESULT	5		
W73476			
ID	W73476	standard; Protein: 2639 AA.	
XX	AC	W73476;	
XX	AC		
DT	29-MAR-1999	(first entry)	
XX		Grapevine leafroll virus type 2 (GRLaV-2) polyprotein.	
DE			
XX		GRLaV-2; closteroVirus; grape; tobacco; transgenic plant;	
KW		disease resistance; virus resistance; beet yellows virus;	
KW		tristeza virus; protease; methyltransferase; helicase.	
XX			
OS		Grapevine leafroll virus type 2.	
XX			
PN		WO9853055-A1.	
XX			
PD		26-NOV-1998.	
XX			
PF		20-MAY-1998; 98WO-US10313.	
XX			
PR		20-MAY-1997; 97US-0047194.	
XX			
PA		(CORR ) CORNELL RES FOUND INC.	
XX			
PI		Gonsalves D, Ling K, Zhu H;	
XX			
XX		WPI: 1999-045307/04.	
DR		N-PSDB; V08864 and V08874.	
DR			
XX			
PT		Grapevine leafroll virus (type 2) proteins and polypeptides - and	
PT		encoding DNA, useful e.g. to impart grapevine leafroll resistance to	
PT		grape and tobacco plants and detect grapevine leafroll virus	
XX			
PS		Claim 4; Page 25-33; 151pp; English.	
XX			
CC		This is the amino acid sequence of a 294 kDa polyprotein encoded by	
CC		open reading frame ORF1a (see V08864) of grapevine leafroll virus	
CC		type 2 (GLRaV-2) RNA (see V08874). It contains conserved domains	
CC		characteristic of 2 papain-like proteases, a methyltransferase and	
CC		a helicase. The GLRaV-2 genome includes 9 open reading frames	
CC		(see V08864-72) encoding the polyprotein, an RNA-dependent RNA	
CC		polymerase, heat shock proteins, coat proteins and other proteins	
CC		of unknown function (see W73476-84). These can be used to produce	
CC		antibodies useful to detect GLRaV in samples e.g. by ELISA	
CC		(claimed). The nucleic acid molecules can be used to produce	
CC		probes and primers for such detection, and to transform host cells	
CC		(especially Agrobacterium vitis, Agrobacterium tumefaciens, grape,	
CC		citrus, beet or tobacco cells) and produce transgenic plants	
CC		(claimed). They can be used to impart GLRaV-2 resistance to vitis	
CC		scion or rootstock cultivars or Nicotiana cultivars (claimed).	
CC		Because extensive similarity exists between hsp70-related sequence	
CC		regions of GLRaV-2 and other closteroViruses, the DNA may also be	
CC		used to impart beet yellows virus resistance to beet cultivars or	
CC		tristeza virus resistance to citrus scion cultivar/rootstock	
CC		cultivars (claimed).	





Db 2116 vvddnkdgtgltnavre-----fyymqelalfelhsklctyyqqlrlvnfdrrsvapcseda 2171  
QY 1869 PVLVLAETKGIPTGKQDVDA-VVALGDNLFVPCDIL-----VFHDALNLGALKV 1916  
Db 2172 qlvvrkngdstivggkevrlhikdfhdhdfldfgkisinrrrrggnvlyhd--nl--afila 2227  
QY 1917 ARCGMVGSEF-----KSEFYKCYNAPPGGGKTTTLVDEEV---KSPNSTATITANV 1964  
Db 2228 snlflagypfsrsvftssvdillieappgggktttidsflkvfkkegstmltank 2287  
QY 1965 GSSEDINMAVKRDPNLE-----GLNSATTNRSRVNFIVRGMVKRVL-VDEVHMM 2014  
Db 2288 ssqveilkkeveksniecqkrkdrspkkslytidaylmhbh--rgcdadvlifdecfmv 2345  
QY 2015 HUGLLQLGVFATGASEGLFFGDINOIPINR---EKVFRMDCAVFVPKESVYVTSKSYR 2071  
Db 2346 hagsvlaclfeftchkvmlfgdsrqihyielneldkclgyldrfv-dlqcrvyngnlsyr 2404  
QY 2072 CPLDVCYLLSSM-----TVRGTEKCYPEKVVSGDKPKVRSLSKRPIGTTDDVAEINAD 2125  
Db 2405 cpwdvcawlsvtvgnlratvkgese-----gks-----smrineinsvddlvpdvgs 2451  
QY 2126 VYLCTOLEKSDMKRSLKGG-KETPVMTVHEAOGKTFSDVVLFRTKKADDSLFTKQPHI 2184  
Db 2452 tfclmlqsekfeiskhfrkgltklnvltvheaqgetyvarvnlvrlkqedeppksirhi 2511  
QY 2185 LVGLSRHRSRLVYALSSKLDKVKCTYISDAS 2216  
Db 2512 tvalsrhtdsitynvlaaarrgdatcdaiqkaa 2543

## RESULT 6

Y91969  
ID Y91969 standard; Protein: 309 AA.

AC Y91969;

DT 19-JUL-2000 (first entry)

XX PMaV-2 helicase.

DE Pineapple mealybug wilt virus 2; PMaV-2; ORF 1a; helicase;  
KW transgenic pineapple; resistance; antiviral.

XX Pineapple mealybug wilt virus 2.

OS WO200017372-A2.

PN 30-MAR-2000.

XX 22-SEP-1999; 99WO-US22152.

XX 23-SEP-1998; 98US-0101461.

XX (UYHA-) UNIV HAWAII.

XX (UYFL) UNIV FLORIDA.

XX Hu JS, Karasev AV, Dawson WO, Melzer M;

XX WPI; 2000-283596/24.

XX N-PSDB; A08690.

PT Isolated pineapple mealybug wilt virus proteins and polypeptides,  
PT useful for protecting pineapples against the virus

PS Claim 17; Page 97-98; 112pp; English.

CC Pineapple mealybug wilt virus 2 (PMaV-2), open reading frame (ORF) 1a  
CC is incomplete but encodes this protein which encompasses all motifs  
CC characteristic of viral helicases. The DNA and protein sequences  
CC are useful for production of transgenic pineapple plant cultivars and  
CC also in transformation methods to impart resistance against the virus to  
CC pineapple plants. Mealybug wilt is a major problem limiting profitable

CC pineapple production in many pineapple growing areas worldwide, the  
CC present invention aims to overcome this problem.  
SQ Sequence 309 AA;

Query Match 5.5%; Score 631.5; DB 21; Length 309;  
Best Local Similarity 44.7%; Pred. No. 7.6e-43;  
Matches 136; Conservative 50; Mismatches 99; Indels 19; Gaps 4;

QY 1949 EFVKSPNSTATITANVGSSSEDINMAVKRDP-----NLEGLNS-----ATTVNSR 1993  
Db 6 eyfkarsgscintanrgsaiddintiesidaanaskaasannvsgvesidnyvcartvnsq 65

QY 1994 VYNFIVRGM--YKRVLVDEVHMHQGLLQGVFATGASEGLFFGDINOIPINREKVFMR 2051

Db 66 lmm--ckgvmmnytcailvdemylmhkgllmlgvfssgarraifgdingqpfinkecfys 123

QY 2052 DCAVFPKESVYVTSKSYRCPLDVCYLLSSMTVTRTEKCYPEKVVSGDKPKVRSLSKR 2111

Db 124 kegvyvcpkgdeilytsesyrpcadvcmwssslkagagsnrylkgvscnqrevvrlsksr 183

QY 2112 PIGTTDDVAEINADVVLCTOLEKSDMKRSLKGGKETPVMTVHEAOGKTFSDVVLFRTK 2171

Db 184 pvyaevqilceadayitfkqckekvrralravgrdrkvtfsheaqgmfgrrvlicrls 243

QY 2172 KADDSLFTKQPHILVGLSRHRSRLVYALSSKLDKVKCTYISDASQVSQSDALLHTFAPA 2231

Db 244 atcdsvfssephlvalsrhtqscvyatlsskladkvgaaidsvtrkevsdvtlktfvas 303

QY 2232 GCGR 2235

Db 304 alfr 307

## RESULT 7

Y58143

ID Y58143 standard; Protein: 120 AA.

XX Y58143;

XX 07-MAR-2000 (first entry)

XX GLRaV-3 polyprotein proteinase domain.

XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;  
KW viral disease; yield loss; sugar content; inhibition; infection;  
KW replication; polyprotein; domain; proteinase; methyltransferase;  
KW helicase; RNA-dependent; RNA polymerase; untranslated region;  
KW transgenic plant; component; resistant; truncation; deletion; antisense;  
KW expression; detection; antibody.

OS Grapevine leafroll-associated virus 3.

XX WO9955880-A1.

XX 04-NOV-1999.

XX 29-APR-1999; 99WO-US09307.

XX 29-APR-1998; 98US-0083404.

XX (CORR ) CORNELL RES FOUND INC.

XX Gonsalves D, Ling K;

XX WPI; 2000-062035/05.

XX N-PSDB; 249203.

PT Newly isolated grapevine leafroll virus protein or polypeptide useful  
PT for producing transgenic plants conferring viral disease resistance -

XX Claim 4; Fig 6; 84pp; English.

XX This sequence represents the proteinase domain of the  
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3  
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever  
CC grapes are grown. Although not lethal, it causes yield losses and  
CC reduction in sugar content. The virus encodes several proteins,  
CC which may serve as targets for the inhibition of viral infection or  
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),  
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase  
CC domain (Y58143), a methyltransferase domain (Y58144), and a helicase  
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF  
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.  
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',  
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are  
CC useful for the generation of transgenic plants and plant components.  
CC Such transgenic plants may be resistant to viral disease, for example,  
CC encoding truncated or internally deleted proteins, or via the use of  
CC antisense nucleotides to inhibit viral gene expression. The nucleotides  
CC may additionally be used in the detection of viral nucleic acids in a  
CC tissue sample. The proteins can be used to generate antibodies which  
CC may be used to detect GLRaV-3 in plant samples. The isolation of  
CC GLRaV-3 facilitates the production of agents that reduce the risk of  
CC infection or damage by the virus in vineyards.

XX SQ Sequence 120 AA;

Query Match 5.3%; Score 605; DB 21; Length 120;

Best Local Similarity 100.0%; Pred. No. 1.9e-41;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VSGVSALRGDGKVKLMELARTSTSDVSDFDVVFVAVSNALLVHVRVVPYAPVKRE 144

|||||  
1 vsgvsalrgdgkvvkmlartstsdvsdfdvvfavsnallvhrvrvpyapvkre 60

QY 145 QPKFVAVKDEQKPKRQASHWAVKPTAGVHVPLPKQEALEPAQSVQQSLEEKAAITFG 204

|||||  
61 qpkpavqdcqkprqashwavkptagvvhvplpkqeaalepaqsvpqgsleekaaiftg 120

RESULT 8

Y91963

ID Y91963 standard; Protein: 598 AA.

AC Y91963;

XX Y91963;

XX 19-JUL-2000 (first entry)

XX PMWav-1 helicase.

XX Pineapple mealybug wilt virus 1; PMWav-1; ORF 1a; helicase;

XX transgenic pineapple; resistance; antiviral.

XX Pineapple mealybug wilt virus 1.

OS WO200017372-A2.

XX 30-MAR-2000.

XX 22-SEP-1999; 99WO-0522152.

XX 23-SEP-1998; 98US-0101461.

XX (UYHA-) UNIV HAWAII.

XX (UYFL) UNIV FLORIDA.

XX Hu JS, Karasev AV, Dawson WO, Melzer M;

XX WPI: 2000-283596/24.

XX N-PSDB; A08684.

XX Isolated pineapple mealybug wilt virus proteins and polypeptides,

XX PT

PT useful for protecting pineapples against the virus

XX Claim 5; Page 79-82; 112pp; English.

XX Pineapple mealybug wilt virus 1 (PMWav-1), open reading frame (ORF) 1a

CC encodes this protein with all 8 motifs conserved in the so-called viral

CC helicases. The DNA sequences and related proteins are useful for

CC production of transgenic pineapple plant cultivars and also in

CC transformation methods to impart resistance against the virus to

CC pineapple plants. Mealybug wilt is a major problem limiting profitable

CC pineapple production in many pineapple growing areas worldwide, the

CC present invention aims to overcome this problem.

XX SQ Sequence 598 AA;

Query Match 2.8%; Score 323; DB 21; Length 598;

Best Local Similarity 26.7%; Pred. No. 2.9e-17;

Matches 131; Conservative 86; Mismatches 175; Indels 98; Gaps 22;

QY 1787 FEVVKT-----GTSEDAAVEYLKYLAIIGIERTYRALLMARNIAVTTAEGVLKVPNQVYE 1840

Db 117 fevktlilantlgfsgftgmsfldi-----esytflvr-----rvsgdgyavleach 166

QY 1841 SLPGFH-----VYKSGTDLIEH-----STQDGLRVRLDLP-----YVL 1872

Db 167 slhglraevdvlkvkqeiqlghkhelfcknvqeldrvkykpdgfhynveisniyl 226

QY 1873 I-AKGGITFTKGDVADVVALGDNLFVCDLILVFDH-AINLIGAL-----KVARGCMVG 1923

Db 227 vntkhlyvnn-----dpvvkesegivllep-----heistfnlrsialdlilvkvs----ve 274

QY 1924 ESEKFE-YKCYNAPPGGGKTTTLVDFV-----KSPNSTATITANVGSSEDIN-MAVK 1975

Db 275 evnkaidnvkfnavpvgagkyqikqmrwfdsekddgsallvltscnsadtikafaqe 334

QY 1976 KRDPNLEGNLSATTNSRVNFIVRG-----MYKRVLVDVDEHMHQGLLQGLGVFATGASEG 2031

Db 335 kr---lgkmiqiltvdafll--fqargrnrvlyktilidccymthagilrgiaavkpeec 389

QY 2032 LFFGDINOIPINREKVFPMDCAVFVPKESVVTYSKYRCPDLVCVYLLSSMTVVRGTGEC 2091

Db 390 vlygdrqvfpfnrklindnksflkpslgnysenlitrpcadicwrmsn--vnnngkk- 446

QY 2092 YPEKVVSG-----KDKPVVBSLSKRPITGTTDDVAEINADVYLCMTOLEKSDMKRSLKCK 2145

Db 447 -gdriysqpvklftqskpvlksvtckaafskgdhnlfsqvdvrmftgkneiliseymnr 505

QY 2146 G-----KETPMTVHEAAGKTFSDDVVLFRTRKKADDSLFTKPHILYGLSRHTSLVYAA 2199

Db 506 gigtigdaaktligtvaesggetykrvhlvafkptddqvssmhrvlrshtslslyqfc 565

QY 2200 LSSKLDKVG 2209

Db 566 ipnkmkgig 575

RESULT 9

R05107

ID R05107 standard; protein: 1128 AA.

XX R05107;

XX R05107;

XX 05-OCT-1990 (first entry)

XX Sequence encoded by RNA1 of AMV.

XX Alfalfa mosaic virus; AMV; cryptic infection; probe.

XX Nicotiana tabacum L "Xanthi-nc".

XX US4921802-A.

XX



```
Db 949 psfvrhkhltgkverklit---wrsadatycl-----ekfyknkpvkt 992
Qy 2104 ---VVRSLSKRPICGTDDVABINADVLCMTQLEKSDMKRSLKGGKRPVMTVHEAQGK 2160
Db 993 nsrvlsiovpinspsverntnalytchtqaekavlkqethlkgcdn-ifttheagqk 1051
Qy 2161 TFSDWLFRTRKKADSLFT-KQP-----HILVGLSRHTRSLVYAALSKLDDKYGTYISD 2214
Db 1052 tfdnvyferlertsittdrpingpchgivalsrhkktkfyftiahdsddviynacd 1111
Qy 2215 ASPOSVSIDALL 2225
Db 1112 ag--ntddsil 1120

RESULT 10
R60608
ID R60608 standard; Protein; 1116 AA.
XX AC R60608;
XX XX
XX DT 21-JUN-1995 (first entry)
XX XX
XX DE Tobamovirus replication enzyme subunit.
XX XX
XX KW Tobamovirus; tobacco mosaic virus; vector; transcription; promoter;
XX KW terminator; replication enzyme; replication site; ori; coat protein.
XX OS Tobacco mosaic virus.
XX XX
XX PN JP06233688-A.
XX XX
XX PD 23-AUG-1994.
XX XX
XX PF 29-NOV-1993; 93JP-0297704.
XX XX
XX PR 01-DEC-1992; 92JP-0322100.
XX XX
XX PA (SANY ) SANKYO CO LTD.
XX XX
XX DR WPT; 1994-319190/40.
XX DR N-PSDB; Q71546.
XX XX
XX PT Improved transcriptional vector of plant virus - esp. vectors
XX PT pL5013 and pL5020 which are prep. and mass produced, then
XX PT tested for stability and function
XX XX
XX PS Disclosure; Page 17-19; 36pp; Japanese.
XX XX
XX CC A transcriptional vector comprises in a 5'-3' direction (i) a 5'
XX CC non-transcribed region of tobamovirus, (ii) a replicative enzyme of
XX CC tobamovirus, (iii) a foreign gene connecting region, and (iv) a 3'
XX CC non-transcribed region of tobamovirus. The vector also has a
XX CC transcriptional promoter region upstream of the tobamovirus
XX CC region and a transcriptional terminator region downstream of the
XX CC tobamovirus region, neither of which is recognised by the host cell,
XX CC a replication initiation site which functions in the host cell, and
XX CC a selective marker gene. This sequence is a subunit of the
XX CC tobamovirus replication enzyme described (See also R60609) and the
XX CC sequence encoding it was used in a construct (See Q71546) which in
XX CC turn was used in the construction of the transcriptional vectors
XX CC pL5013 and pL5020.
XX XX
XX SQ Sequence 1116 AA;

Query Match 2.7%; Score 309; DB 15; Length 1116;
Best Local Similarity 16.8%; Pred. No. 1.2e-15;
Matches 315; Conservative 205; Mismatches 519; Indels 840; Gaps 64;

Qy 403 ALNREGEVSSVVVITNRLVLRDQSALLSHLDTKLCDMFSQORDAMIREKPSHRCDVFLKP 462
Db 2 aylqtatssalletvrgmntlrvldlakrrlydtav-defnard-----rrpkvnfskvvse 56
```

```
Qy 463 REREKRLRFPPELSIQFSDSVSRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
Db 57 eqtliatkaypefgitfyntqnavhsaggirslrleyleymmqipysltydignfashl 116
Qy 522 KAGHVNCVHCNPVLVDKVDKRRINE----ILFLS-TAGGDSYVS-----SDLTE 566
Db 117 fkgayvhccmpnldvrdimrheggkdsielylsrlergnkhvnpfqkeafdryaempne 176
Qy 567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQOVAEAMDKKALYDFDIALMPPVELL 625
Db 177 vvchdtfqtrhsqecygrvyaialhsyidpadefgaalrknvhnvcyaafhsenll 236
Qy 626 YGNGEVLLEELDTLVKREGDYLAENVGCGEM-YEHSEFSNVSGFFTSYVTSNGNVFKL 684
Db 237 ledshvnaldeinaacqrdgdrltfsfasestlnyshsnylnkkycktyfpasrevymk 296
Qy 685 EYEGYRCGYHHLTMCRQAOKSPGTEVTVRSVPSFVGKSLVPIPVVAGSSVSFKTIVLDS 744
Db 297 ef----- 298
Qy 745 FVDRIYSY--ALNTIGTFENRTFEYAVCAVRSQKTHVITGSRVVSXKVDISPDMMWGLVV 802
Db 299 lvtrvntwfcfsridtf-----llykgvahkgvd----- 328
Qy 803 AVMAQAIDRAKRSIRSYNFIKASEGSLAGVFKLFFQVDCFSNAVSYYAKAMVHDNFNV 862
Db 329 -----seqfykamedawh-----ykktlamcnseriliedssvnywfpk 368
Qy 863 LETLMSPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRLKLNRLRVFSRA 922
Db 369 mrdmviwp-----lfdislet-----sktrkevl--- 393
Qy 923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCSTRAGLLGGSATVVSYSVK 981
Db 394 vskdfvtylvnhirtyakal--tysnvlsvfvesirsv-----ilngvtarsewdvdk 445
Qy 982 GLVARGAATKAFSGITSPFSTGSLFYDRGLTEDERLDALVRTENAINSPGILETSRVAV 1041
Db 446 sliqslsmt-----ff-----lhcklavlkdd-lli 470
Qy 1042 SKVVAGTK-----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRIAANARR 1097
Db 471 skfalgpktvsqhwdeisl----- 490
Qy 1098 YAGSSYETLSSLSQAAGLRLTSSITVSGSLVVRRCFSSAVTVTRATVAKRQVPLALL 1157
Db 491 -----afgnafpsikerlinrk----- 507
Qy 1158 SFSTSYAISGCSMLGIWAHALPRHLMFPFGLTLLGARASANTWKFGGFSNNCAVPE-- 1215
Db 508 -----likitenale-----irvpdly 524
Qy 1216 VVWRKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNYYDETLYRYRDFDYDE 1275
Db 525 vtfndrlvseykmsvdm-----pvlidrkmeeteemy----- 557
Qy 1276 GAGSGTQHEAVPGDDNDGCTSSVSSYDVTVNVRDVGISTNGEVTGEEETHSPRSVQVTY 1335
Db 558 -----nalselsvlnksdkfdvdfsqmc-----qsl----- 584
Qy 1336 VEEEVAPSAVAERQGDPSGSGTADAMAFVESVKKGVDVDFHQOSSGSETAREVEVDGKGL 1395
Db 585 ---evdpmta-----akvivavmsnesgiltfdeqteanvalal----- 621
Qy 1396 LPESVWGEPATQERGRAADGNHTAQTAVNEGDRPEVQSSLSVSSPQADIPKVTQSEVHAQKE 1455
Db 622 -----qdsekaadg-----alvlsrdrveepsikgs-----m 648
Qy 1456 VKQEVPLATVSGATPIVDEKPAFVTTTRGVKLIIDGKAVAHAEKKQVQVQOPQRSUTI 1515
Db 649 argelqlaglsq-----dvpessyt-----rseeslesleqfhnat 683
```

```
QY 1516 NEGKAGKOLCMFRTCSGVLQDLDVYNEATITATRESNAFTVDNLKGRSAVFESKLGEGYTY 1575
Db 684 aasllhkmcsi-----vycgplkvqgmkn---fidslvaslsa----- 719
QY 1576 NGGSHVSSGWPRALDILTAIKYPSVFDHCLVQKYMGGVFPFHADDEECYPSDNPILTV 1635
Db 720 -----avsnlvklkdaaid--letqkfg----- 743
QY 1636 NLVGRANFTKCRKGGKGVWVNVASGDYFLMPC-----GFQTHLHSVN-----SIDGR 1685
Db 744 -----vidvasrwlvpksaknhawgvvetharkyhvallehde-- 782
QY 1686 ISLTRATRRVFGVGRMLQLAGGVSDKSPGVNPOQOSOGATRTITPKSGCKALSEGSG 1745
Db 783 -----fqi----- 785
QY 1746 REVKGRSTYSIWCEODYVRKCEWLADNPVMALEPDYTPMTPEVVKYTGTSDAVVEYLKY 1805
Db 786 -----itcd-----nwrr-----vavssesvv---y 803
QY 1806 LAIGIERTYRALLMARNIAVTAEGVLKVPNOVYESLSPGFHVYKSGTDLIFHSTODGLRV 1865
Db 804 sdmaklrtrllrkqgephvsakvvl----- 830
QY 1866 RDLPLYVLAERKGIPTKGKDQDAVVALGDNLFVCDLILVFDHDAINLIGALKVARCGWGES 1925
Db 831 -----vdgv----- 834
QY 1926 FKSFEYKYNAPPGGKTTTLVDFEVPKSPNSTATITANVGSESDINMA-----VKKRD 1978
Db 835 -----pgcgk-----tkellsrvnfeedlilvpgrqaaemirr 868
QY 1979 PNLEGLNATTVNSRW--NFIV-----RGMVKRVLVDEVHMHOGLLOLGVFATGASEG 2031
Db 869 anasgilvatkdnvrtvdsflmnykgarqcqkrlifdeglmhtgcvnflvemsldcia 928
QY 2032 LFFGDIINOIPFINREKVFEMCAVEFPKESVYTSKSYRCPLDYCYLLSS-----MTV 2085
Db 929 yvvgtdgplpylnrvtgifypahfaklevdevetrtrtlrcpadvthlndryeghvmct 988
QY 2086 RGTEKCYEKKVVG--KDKPVVRSLSKRPIGTDDVAEINADVYLCMTOLEKSDMKRSLK 2143
Db 989 ssektsvsgemvsgaaslnpyskplkgk-----iltftq---sd-keall 1029
QY 2144 GKGETPVMTVHEAGKTFSDVFLFRTKKADDSLFK--OPHILVGLSHRHTSLVYAA--- 2199
Db 1030 srg-yadvhtvhevgetyadvslvrlcptvpsiiardsphvlvshtsklskyyttvm 1088
QY 2200 --LSSKLLDD--KVGTYISD 2214
Db 1089 dplvsilrdlervasyllid 1107

RESULT 11
W87724
ID W87724 standard; Protein; 2161 AA.
XX
AC W87724;
XX
DT 26-APR-1999 (first entry)
XX
DE Rupestris stem pitting associated virus RSPav-1 replicase.
XX
KW RSPav-1; grape; transgenic plant; disease resistance; replicase.
XX
OS Rupestris stem pitting associated virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 1931
FT /note= "encoded by AG, apparent 1 nucleotide
FT deletion in open reading frame"
PN WO9852964-A1.
```

```
XX 26-NOV-1998.
PD
XX
PF 20-MAY-1998; 98WO-US10391.
XX
PR 17-DEC-1997; 97US-0069902.
PR 20-MAY-1997; 97US-0047147.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Meng B;
XX
XX WPI; 1999-045297/04.
XX N-PSDB; V99284.
XX
XX Isolated proteins from Rupestris stem pitting-associated virus and
XX related nucleic acid - vectors, host cells and transgenic Vitis
XX cultivars that are resistant to the virus
XX
XX Claim 4; Page 18-25; 163pp; English.
XX
XX This is the amino acid sequence of the replicase of Rupestris stem
XX pitting associated virus RSPav-1. It was deduced from the
XX nucleotide sequence of an open reading frame (ORF1) identified
XX in the RSPav-1 genome (see V99284). Proteins and polypeptides
XX of RSPav (see W87724-38), and nucleic acids encoding them (see
XX V99284-86) are claimed, as are expression systems, transformed
XX host cells, transgenic grape (Vitis) scion or rootstock cultivars
XX comprising DNA encoding a replicase, coat protein, or a protein
XX of a triple gene block, and a method of imparting RSPav resistance
XX to Vitis scion or rootstock cultivars by transformation with such
XX DNA.
XX
XX Sequence 2161 AA;
XX
XX
XX Query Match 2.5%; Score 290; DB 20; Length 2161;
XX Best Local Similarity 41.5%; Pred. NO. 1.4e-13;
XX Matches 59; Conservative 26; Mismatches 53; Indels 4; Gaps 3;
XX
QY 1554 FVDNLKGRSAVFESKLG-EGYTYNGSGVSHVSSGWPRALDILTAIKYPSVFDHCLVQKYM 1612
Db 733 fxdrlrgasffskpgischsngsyngstsgwpmqqlsstggrnyynsclaqiye 792
QY 1613 GGGVFPFHADDEECYPSDNPILTVNLVGRANFS--TKCRK--GGKVMVNVASGDYFLMPCG 1669
Db 793 nsklahhkddescyeighkvltvnligsatftiskrnlvggnhcsltigpnneffemprg 852
QY 1670 FQTHLHSVNSIDEGRISLTFR 1691
Db 853 mqcnfghvsnctpgrvsltr 874

RESULT 12
R06531
ID R06531 standard; protein; 992 AA.
XX
AC R06531;
XX
DT 19-DEC-1990 (first entry)
XX
DE Protein encoded by RNA 1 gene of CMV.
XX
KW CMV; RNA 1.
XX
OS Cucumber mosaic virus.
XX
PN JP02167080-A.
XX
PD 27-JUN-1990.
XX
PF 19-DEC-1988; 88JP-0320015.
XX
```

```
PR 19-DEC-1988; 88JP-0320015.
XX (NOGY-) NOGYO SEIBUTSU IDEN.
XX WPI; 1990-241950/32.
DR N-PSDB; Q05751.
XX Genome RNA 1 gene for prepn. of cucumber mosaic virus - codes
XX of resistant plants.
XX Claim 1; Page 1; 10pp; Japanese.
PS The sequence was deduced from the DNA sequence which was prepd.
CC from RNA isolated from CMV-O virus. The gene can be used to
CC prepare plants which are resistant to cucumber mosaic virus.
XX Sequence 992 AA;

Query Match 2.4%; Score 274.5; DB 11; Length 992;
Best Local Similarity 22.2%; Pred. No. 6.3e-13;
Matches 146; Conservative 101; Mismatches 267; Indels 145; Gaps 25;

QY 1566 FSKLGEYTYNGSHVSSGWPRALEDILTATKYPVSFVDFHCLVQYKMGGGVFPFHADDEEC 1625
DB 434 yeqlgkmydmwnassisk-w-----faaltrrrvffssav-----halfptl 474

QY 1626 YPSNPIILTVNLCKANFSTKCR-KGCKVMVINVASGDYFLMPCGFORTHLSVNSNIDEG 1684
DB 475 rpreekeflkistvfneecsfdggeewdv-issaayv-----atqavtdg 521

QY 1685 RISLTERATRVFCVGRMLQLAGVDEKSPVNPQOQSGATRTITPKSGGKALSEG 1744
DB 522 kvlaaqkaekl---aeklagpvdevd--spevpsstd----- 555

QY 1745 GREVKGRSTYSIMCEQDYVRKCEWLADNPVMALEPDYTPMTFEVVKTGTSDEAVVEYLK 1804
DB 556 -----dtadvcgekevsceldlsaq-----rspit-rvaerat---amleyaa 596

QY 1805 YLAIGIERTRALLMARNIAVTTAEGVLKVPNQV-----YESLPG-----PHVYKSGTDL 1854
DB 597 y-----ekqlhdtvtvsnlkriwnmagggdkrnslegnlkfvdtyftvdpm 642

QY 1855 IFHSTQDGLRVRLPYVLLAEPKGFITKG---KDVDVAVVALGDNLFCVDDILVFFHDAINLI 1911
DB 643 vnlfstgrvmrpydegivysvgynerglpgksdgelifvnscevicns-----eslsav 697

QY 1912 GALKVARCGMVGESFKFEYKCYNAPPGGGKTTTLVDDEFVKSPNSTATITANVGSSSEDIN 1971
DB 698 trslqaptgtisq-----vdgvagcgktttaiksf--epstdmivtanksaqdvr 746

QY 1972 MAVKKRDPNLEGLNSATVNSRVNFIVRGMKRVLVDEVHMHQGLLQLGVFATGASEG 2031
DB 747 malfkssdskeacafvrtadsvllnec--ptvsrvlvdevvllhfgqlcavmsklkavra 804

QY 2032 LFFGDINQIPINREKRVMDCAVFPVKESVVTYSYRCPLDVCPLLSMTV-----RG 2087
DB 805 icfgdsqiafsrdsafdmrfskiiidetsdaet--tfrrpqdvplvrlmactaklpkg 862

QY 2088 TEKCYPEKVVSGDKPVVRSLSKRPIGTTDDVAEINADVLYCMTQLEKSDMKRSLKGGK 2147
DB 863 thskytkwvsgsk---vkrsvtssiasvtlvdldssrfyitmtqadka----slisrak 915

QY 2148 E-----TPVMTVHEAOGKTESDVLPFRKKADSLFTKQPHILVGLSRHTRSLVY 2197
DB 916 emnlpkftwnerikthvesgdgisedhvtlvrkstkcdlrfkqfscylvaltrhkvrtiry 974

RESULT 13
R49657
ID R49657 standard; Protein; 1704 AA.
XX
```

```
AC R49657;
XX 12-SEP-1994 (first entry)
XX Sequence of Heliothis armigera replicase encoded by RNA 1.
DE HasV; RNA 1; small RNA virus; replicase.
XX Heliothis armigera stunt virus.
XX WO9404660-A.
XX 03-MAR-1994.
XX 13-AUG-1993; 93WO-AU00411.
XX 14-AUG-1992; 92AU-0004081.
XX 08-JUL-1993; 93US-0089372.
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX (PACI-) PACIFIC SEEDS PTY LTD.
XX Christian PD, Gordon KHJ, Hanzlik TN;
XX WPI; 1994-083180/10.
XX N-PSDB; Q58522.
XX Small RNA virus capable of infecting insect species, e.g.
XX Heliothis - and transgenic plants contg. viral nucleic acid, for
XX protection against insect pests
XX Disclosure; Figure 1; 183pp; English.
XX The inventors claim a virus comprising a genome hybridisable with
XX the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
XX those given in Figs 1 and 2 of the specification. Isolated proteins
XX or polypeptide prep. of the proteins or polypeptides derivable
XX from the virus are also claimed.
XX H. armigera larvae were raised and viral RNA was extracted. The virus
XX RNAs were reverse transcribed into cDNA. Clone E3 represents 99.7%
XX of RNA 1.(hr236 contains about 88% or RNA 2.)The full length clone
XX of RNA 1 was completed using PCR. RNA 1 encodes a protein of mol.
XX wt. 187,000 which is regarded as the replicase in view of its AA
XX sequence similarity in certain limited regions to replicases of
XX of other RNA viruses. The apparent mol. wt. of this protein upon
XX in vitro translation of virus RNA and SDS-PAGE is 195,000. The
XX sequence given in Figure 1 is in the same sense as the viral
XX (positive-sense) RNA. There are other small ORFs at the 3' end,
XX corresp. to the proteins P1la, P1lb and P14.
XX Sequence 1704 AA;

Query Match 2.0%; Score 233; DB 15; Length 1704;
Best Local Similarity 20.9%; Pred. No. 3.9e-09;
Matches 206; Conservative 108; Mismatches 330; Indels 342; Gaps 47;

QY 1425 GDREPVQSSLVSSQADIPKVTQSEVHAQKEVQEVLATVSGATPIVDEKPAFVTRG 1484
DB 54 gtlpbtqphllaghqrvaeevlhmfargrstvlelgr--slhsalklhgapnapeadyhg 111
QY 1485 VKIIDKGKAVAHVA--EKKQVQVEQPK---QRLTINEGKAGKQLCMFRTCSG-----G 1533
DB 112 ctkygtgrdgrsrhialeersvatgrpefkadasllan-gigrftfcvvgvsgcafsrvg 170
QY 1534 VOLDVYNEATATFRSNAFTFDNLKGRSAVFFSKLOGEYTYNGSHVSSGWPRALEDIL 1593
DB 171 ianhslydvtl-eelanafe-----nhghlmvrfrfmppeall 207
QY 1594 -----TAKYPSVFDHCLVQYKMGGGVFPFHADDEECYPSDNPIILTVNLVG 1639
DB 208 ymdnvvnaelgyrfhvieepmavkdcafq----ggdlrlhfpeidfinesqerrierlaa 263
```

```
QY 1640 KANFSTCKRKCKGMVINVASGD-----YFLM-----PCGF-----OR 1672
Db 264 rgsyrra-----vifsdddgdgdaylhdftwtwpyllvrnyptbfgfslhievqr 314
QY 1673 THLSVNSIDEGRI SLTFRATRVRFGVGRMLQLAGVSDKSPGVPNQOQSQGATRTIT 1732
Db 315 rhgssie-----lritrappg-dcmilav-----vpr-----tsqglcr--i 347
QY 1733 PKSGGKALSEGSGREVKGSTYSINCEQ-----DVRKCEMLRADN 1773
Db 348 pnifyya--dasgtehk-----tiltsqhkvmllnfmqtrpekelvdmvlnmsfarar1 400
QY 1774 PYMALEPDYTPMTFEVVKVTGTSEDAAVEYLKYLAIIGIERTYRALLMARNIAVTTAEGVLK 1833
Db 401 raivvasevteswnispadlvrtvvslyvlihi--ierrra-----vavktakdvf 451
QY 1834 VPNOVYESLPGPHVYKS-----GTDLIF-HSTQDGLVRDLPYVL-----1872
Db 452 getstweslk--hvfsgccglrnlkgtvdfkrvdkrvhslgdilcdvrlspeqvqf 509
QY 1873 -----IAEKGIPT-----KGKDVD-----A 1887
Db 510 lpsrvpparvfhdrealealreagcynerpvpstptpveepqgfdadlwhataasipeyra 569
QY 1888 VVALGDNLFVCCDDILVFHDAINLICALKVARC-GMVGESFKFEYKCYKNAPPGGCKTTTL 1946
Db 570 tiqaglntdvklkltlenalktidgltisvrgl-----emyegppgsaktgtl 619
QY 1947 V-----DEFVKS PNSTATITANVGSSSEDINNAVKKRPNNLEG 1983
Db 620 iaaleaaggkalyvaptrelreamdrrikppsasat-----qhvvalalirr-ataeg 670
QY 1984 LNSATVNSRVNFTVRGMYKRVLVDEVHMHQGLQLGVFATGASEGLFFGGINOIPPI 2043
Db 671 apfatvvidecf-----mfplvyvalvhal-----spssrivilgdvqhigfi 713
QY 2044 NREKVFWRMDCAVFPKESVYVTSKYRCLP--DVCYLLSSMTVRCTEKCYPEKVVSXGD 2101
Db 714 d-----fqqtsaamplvrdvkvqrrfntqtkrc-padvva---749
QY 2102 KPVRSLSKRPIGTTDDVAEIN--ADVY-----LCMTQLEKSDMKRSLKKGKRETPVM 2152
Db 750 ttfifsglypgcttgscvasishvapdrnsqaqlclftqeeks--rhgaeg-----am 801
QY 2153 TVHEAQGTFSQVVL-FRTKKADDSLFKQPHILVGLSRHSRLSYAALSSKLDKVGTY 2211
Db 802 tvheaqgrtfasviihyngstaeqkllaekshllvgitrthnl-----y 846
QY 2212 ISDASQSVSDALLHTFAPAGCFRCI 2237
Db 847 lrd--ptgdierqlnhsakaevftdi 870

RESULT 14
W56322
ID W56322 standard; Protein; 2039 AA.
XX
AC
XX
XX
DT
XX
DE Haemophilus paragallinarum antigenic protein #2.
XX
KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW vaccine; chicken infectious coryza; CIC; fowl.
XX
OS Haemophilus paragallinarum.
XX
FH Key Location/Qualifiers
FT Peptide 1..70
FT Protein /label= signal
71..2039
/note= "antigenic protein"
```

```
XX WO9812331-AL.
PN
XX
XX 26-MAR-1998.
XX
PF 12-SEP-1997; 97WO-JP03222.
XX
PR 19-SEP-1996; 96JP-0271408.
XX
PA (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
PA (KAGA ) CHEMA-SERO-THERAPEUTIC RES INST.
XX
PI Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX
DR WPI; 1998-230318/20.
DR N-PSDB; V22837.
XX
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
XX antibody production - and is useful for diagnosis of and preparation
XX of vaccines for chicken infectious coryza
XX
PS Claim 5; Page 71-87; 108pp; Japanese.
XX
XX The present sequence represents an antigenic protein derived from
XX Haemophilus paragallinarum strain C-53-47. The antigenic protein
XX stimulates the production of HI antibodies in fowl. The protein
XX and DNA coding for it can be used in the preparation of vaccines
XX for the prevention of chicken infectious coryza (CIC). The protein
XX and its antibodies can be used in the diagnosis and treatment of CIC.
XX
SQ Sequence 2039 AA;

Query Match 1.6%; Score 185; DB 19; Length 2039;
Best Local Similarity 20.3%; Pred. No. 4.3e-05;
Matches 263; Conservative 147; Mismatches 495; Indels 388; Gaps 61;

QY 920 SRAIVEDSTKVMKAKTEGKPLPITEDSVYAFIMGNVSNVHCTRAGLGGSKATVVSSV 979
Db 380 srtfakdsval--gnkteasn---agsmayg-----kakavgagalaigaev 422
QY 980 SKG-----LVARGA-AT-----KAFSGITSPFSGSLFYDRGLTEDERL 1017
Db 423 aagaefdsdqagllnrgayatlksadskddikagdainvf-----tgfdmmltgqshl 478
QY 1018 DALVRTEANA--INSPVGILETSRVAV-----SKVWAGTKEFWSEVSLNDFTFVLRNKVL 1070
Db 479 ----tyenttyittsagdikktlaavdgogknaiaignkktfaska-----nsva 523
QY 1071 IGTFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSLSO-----AAGGLRGLTSSTVS 1126
Db 524 lg-----syalasagna--falgsylsvsplaantivlgvgyatgsnsfv- 567
QY 1127 GCSLVVRGFSsavivtravakrovpallallsfsysaiysgcsmclgiwahalprhlmpff 1186
Db 568 ggsww-----stlsartv---vlgysasissdshslamgvna-----f1 604
QY 1187 GLGTLGARASANTWKGFGFSNNWCAVPEVWVRGKSVSSLLPLITLIGVSLIIRGLLNDTI 1246
Db 605 gng-----snssialgtgstiaknaks---pdsiaigkdsridakdcd 644
QY 1247 POLAYVPVPEGRNVYDETLRYRYDFDYDE-----1275
Db 645 ngvlytp-----qvydettrafttdenkdyrmqamalgnakvsgkqkmetglnsmal 699
QY 1276 GAGPSGT-----OHEA-----VPGDDNDG---STSSVSVDV 1305
Db 700 garsqatqlnstalgnaktdytwegleapwvskgaisiptsgkigvisvsgksgerri 759
QY 1306 TNVRDVGISTNGEVTGEETHSPR-----SVQYTYVEEB-----VAPSAVAER 1349
Db 760 vnvasgsldtdavnvaqiktieerfqseidlqlngggvgdylsvektningeagrvasqlr 819
```





Search completed: January 31, 2001, 15:29:28  
Job time: 217 sec

---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:26:27 ; Search time 27.64 Seconds  
(without alignments)  
1453.325 Million cell updates/sec

Title: US-09-301-906-15  
Perfect score: 11520  
Sequence: 1 MDYIRPLRVFSPHVNNTLE.....QVSDDLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5247	45.5	1390	2	US-08-770-544-2
2	290	2.5	2161	3	US-09-081-320-3
3	179.5	1.6	1693	3	US-08-840-316-1
4	179.5	1.6	1693	4	PCT-US93-08849A-1
5	179.5	1.6	1693	3	PCT-US93-08849-1
6	169.5	1.5	1693	3	US-08-478-507-7
7	150.5	1.3	1222	2	US-08-682-517-15
8	150.5	1.3	1252	2	US-08-682-517-9
9	149.5	1.3	1794	5	5183745-6
10	142.5	1.2	1489	5	5183745-2
11	142.5	1.2	1848	4	PCT-US95-10661A-6
12	138.5	1.2	1912	1	US-08-409-995-4
13	138.5	1.2	1912	3	US-08-685-467-4
14	135.5	1.2	1302	1	US-08-232-537-2
15	135	1.2	1702	4	PCT-US95-10661A-5
16	134.5	1.2	1541	4	PCT-US95-10661A-3
17	134	1.1	1027	3	US-08-446-1378-2
18	131.5	1.1	2756	1	US-08-375-709-11
19	131.5	1.1	2756	1	US-08-752-929-11
20	131.5	1.1	2756	3	US-09-090-793-7
21	131	1.1	1611	2	US-08-804-227C-5
22	130	1.1	2446	2	US-08-551-356-2
23	130	1.1	2446	4	PCT-US93-12687-2
24	129	1.1	4545	2	US-08-804-227C-14
25	129	1.1	4550	2	US-08-804-227C-8
26	129	1.1	4550	2	US-08-804-198-2
27	127.5	1.1	1780	1	US-08-769-309A-5
28	127.5	1.1	1780	3	US-08-994-570-5

29 127 1.1 1026 1 US-08-194-290-7 Sequence 7, Appli  
30 127 1.1 1026 2 US-08-614-377A-7 Sequence 7, Appli  
31 125 1.1 2324 1 US-08-283-857-1 Sequence 1, Appli  
32 125 1.1 2324 4 PCT-US95-09819-1 Sequence 1, Appli  
33 125 1.1 2327 5 5455158-1 Patent No. 5455158  
34 124.5 1.1 1612 1 US-08-169-927-2 Sequence 2, Appli  
35 124.5 1.1 4472 2 US-08-804-227C-2 Sequence 2, Appli  
36 124 1.1 1529 2 US-08-728-470-10 Sequence 10, Appli  
37 124 1.1 1651 1 US-08-447-411-2 Sequence 2, Appli  
38 124 1.1 2431 1 US-07-920-281C-2 Sequence 2, Appli  
39 123.5 1.1 1069 1 US-07-777-715-9 Sequence 9, Appli  
40 123.5 1.1 1069 1 US-08-170-126-4 Sequence 4, Appli  
41 123.5 1.1 1069 3 US-08-954-418-4 Sequence 4, Appli  
42 123.5 1.1 1545 4 PCT-US95-10661A-4 Patent No. 5180808  
43 123.5 1.1 2409 5 5180808-2 Sequence 2, Appli  
44 123 1.1 628 3 US-08-656-034-2 Sequence 2, Appli  
45 122 1.1 1346 3 US-09-320-878-4 Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-770-544-2  
; Sequence 2, Application US/08770544  
; Patent No. 5907085  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Ling, Kai-Shu  
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND  
; TITLE OF INVENTION: THEIR USES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770,544  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60009008  
; FILING DATE: 21-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/621  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1390 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-770-544-2

Query Match 45.5%; Score 5247; DB 2; Length 1390;  
Best Local Similarity 74.1%; Pred. No. 0;  
Matches 1055; Conservative 96; Mismatches 206; Indels 56; Gaps 17;

Qy 848 VSVYKAMVHDFNVLETLMSPRAFIRKVPGVVVVTICTSGASDRLELRGAFDISKETF 907





```

; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT-US93/08849A
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US07/947,263
; FILING DATE: 18-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: William S. Feller
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4032 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; PCT-US93-08849A-1

Query Match 1.6%; Score 179.5; DB 4; Length 1693;
Best Local Similarity 21.3%; Pred. No. 4.4e-06;
Matches 189; Conservative 98; Mismatches 320; Indels 281; Gaps 45;

Qy 1521 GKQLCMFRTC-----SCGVQLD-----VYNEA-----TIATFRSNAFTFVDNLKGRSAVFF 1566
Db 425 GRQLEFYAQCRRWLSAGFHLDPRLVDFESAPCHCRTAIRKAVSKFCFCFMKWLQECTCF 484

Qy 1567 SKLGEY-----TYNG-----GSHVSSGWPRALEDLITAIKYPVSVD 1603
Db 485 LQPAEGVVGGQHDNEAYEGSDVDPAESALSDISGVTVFC--TALQLYQALDLP----- 539

Qy 1604 HCLVQKYMGGGVFPFHADDECPNDPILTNLVGKANFSTKCRKGKVMVINVASGDY 1663
Db 540 EIVARAGRLATV-----KVSQVDGRIDCETLLGNKTFRTSF-VDGAVLETN----- 585

Qy 1664 FLMPCGQFQTHLH---SVNSIDEGRISLTFRATR-----VFGVRMLQLAGGVSEK 1713
Db 586 -----GPERHNLSPDASQSTMAAGPFSLTUYAASAGLEVRYVAAGLDHRAVFAPGVSPRS 640

Qy 1714 SPGVP-----NQPO-----SOGATRTITPKSGKALSEGSGREVKGRS 1752
Db 641 APGEVTAFCSALYRFNREAOQLSLTGNWFHPEGLLGPFPAPSPGHVWE--SANPFCEG 698

Qy 1753 TY---SIWCEQDYVRKCEW-----LRADNPVMAL-----EPDYTPMTPEVVKTG 1793
Db 699 TLYTRTWSEVDVAPSPAQDLGFTSEPSISRAATPTPAAPLPPADPSP-TLSAPARG 757

Qy 1794 TSEDVAVVEYLYKLAIGITERYAL-----LMARNIYVTTAGVLKVPN----- 1836
Db 758 EPAPGATARAIPAITHQTAHRRLIFTYPDGSKVPAGSLFESTCTWLNVNASNVDRHPGGGL 817

Qy 1837 --QVYESLP-----GFHVYKSGTDLIFHSTQDGLVRDLP----- 1869
Db 818 CHAFYQRYPASFASVMDRDAAYTLTPRPILHAVDPYRLHNPKRLEAAVRETCRSR 877

Qy 1870 -----YVLIAEKGIFFK--GKDVDAVV---ALGDNI--V-----CDDILVFF 1905
Db 878 LGTAAAYPLLG-TGIYQVPIGSPFDAWERNHRPCGDELY.PELAAWFEANRPTCTTLITE 936

Qy 1906 D-----AINLGALKVAC-----CMVGESFSEYKCYKNAPPGGKTTTLVDEF 1950
Db 937 DVARTANLAIELDSATDVGACACRCRTVPGV-----QYQETAGVPGSGKRSI----- 985

Qy 1951 VKSPNSTATTANVGSSDENMAVKKDPNLEGLNSATTVNSRVNFIIVRGMYKRVLVDE 2010
Db 986 -----TOADVDDVVVPTRELNAWRRR-----GFAAPTPTAARVT-----QGRRVVIDE 1030
```

```

Qy 2011 VMMHQGLLQGVFATGASEGLFFCDINQIPFINREKVRMDCAVFPK-RESVVYTS-- 2067
Db 1031 APSLPPHLLHLM--QRAATVHLLGDPNQIPAIDFEH-----AGLVPAIRDPLAPTSWW 1082

Qy 2068 -KSYRCPLDVCYLLSSMTVRGTEKCYPEKVSGKDKPVVRSL--SKRPIGTDDVAEINA 2124
Db 1083 HVTHRCPADVCEL-----IRG---AYPMIQTTSR---VLRSLFWGEPVAGOK----- 1123

Qy 2125 DVYLCMTQLEKSDMKRSILKGGKETPVMTVHEAOGKTFSDVVLFRTKKADDSLFTKOPHI 2184
Db 1124 ---LVFTQAANKAPGS-----VTVHEAOGATTTETIATADARGLIOSSRAHA 1170

Qy 2185 LVGLSRHTRSLVYAAALSSKLDKVGTYISDASPOSVDALLHTFAPAG 2232
Db 1171 IVALTRTEKCVIIDAPGLLRE-VG-----ISDAIVNFFLAG 1207

RESULT 5
PCT-US93-08849-1
; Sequence 1, Application PC/TUS9308849
; GENERAL INFORMATION:
; APPLICANT: Tsarev, Sergei A., Emerson, H.
; APPLICANT: Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08849
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,263
; FILING DATE: 18-SEP-1992
; NAME:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bork, Richard, W.
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; PCT-US93-08849-1
```

```

Query Match 1.6%; Score 179.5; DB 4; Length 1693;
Best Local Similarity 21.3%; Pred. No. 4.4e-06;
Matches 189; Conservative 98; Mismatches 320; Indels 281; Gaps 45;

Qy 1521 GKQLCMFRTC-----SCGVQLD-----VYNEA-----TIATFRSNAFTFVDNLKGRSAVFF 1566
Db 425 GRQLEFYAQCRRWLSAGFHLDPRLVDFESAPCHCRTAIRKAVSKFCFCFMKWLQECTCF 484

Qy 1567 SKLGEY-----TYNG-----GSHVSSGWPRALEDLITAIKYPVSVD 1603
```



Db 699 TLYTRTWSEVDAVSSPARPDLCFMSERPSIPSRAAATPTLAAPLPDPSP-----P 750  
Qy 1794 TSEDVAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNOVYESL----- 1842  
Db 751 PSAPALAEPAASCATAGAPAIHTOTARHRRLLFTTYPDGSKVFAGSFESTCWLNVNASVD 810  
Qy 1843 --PG-----FHVYKSGTDLIHFHSTODG-----LRVRDLPVYLIAEKGIPTKGKDVAV 1888  
Db 811 HRPGGGLCHAFYQRYPASFDASFVMRDGAAYTLTPRPIIHAVAPDYRLHNPKRLEA- 869  
Qy 1889 VALGDNLPVCDIIIVFHDAINLIGALKVARGCM-----VGESPKSFEYKCYNAPPGG-- 1940  
Db 870 -----AYRETC SRLCTAAYPLLCGTCIYQVPIGPSFDABE---RNHRPGDEL 912  
Qy 1941 -----GKTTTLVDEFVKSPNSTATITANVGSSSEDINMAVK----- 1975  
Db 913 YLPELARWFEANRPRTPTLTITEDVAR-----TANLAIELDSATDVGACAGCRVTPGV 968  
Qy 1976 --KRDPNLEGLNSATTVNSRVNFIIV-----RGMV-----KRVLV 2008  
Db 969 QYQFTAGVPGSGKSRISITQADVVDVVVPTRELNRNRRRGFAFTPHTAARVTOGRRVVI 1028  
Qy 2009 DEVHMHQGLLQGVATGASEGLFFGDINOIPEINREKVFMRDCAVEVPK-KESVYVTS 2067  
Db 1029 DEAPSLPHLLHLM--ORAAVHLLGDPNQIPADFEH-----AGLVPAIRPOLGPTS 1080  
Qy 2068 ---KSYRCLPDVCLLSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPTGTTDDVAEI 2122  
Db 1081 WHVHTRWAPDCEU-----IRG---APYMIQTTSR---VLRSFWGEPAVGQK----- 1123  
Qy 2123 NADVYLCMTQLEKSDMKRSLKGGKETPVMTVHEAOGKTFSDVVLFRTKKADDSLFTKQP 2182  
Db 1124 -----LVFTQAOKANPGS-----VTVHEAOGATYTTETIIATADARGLIQSSRA 1168  
Qy 2183 HILVGLSHRTSLVYALSSKLDDKVGTYISDASPQSVSDALLHTFPAG 2232  
Db 1169 HAIVALTRHTEKVICIDAPGLLRE-VG-----ISDAIVNFFFLAG 1207

RESULT 7  
US-08-682-517-15  
; Sequence 15, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/682,517  
; CLASSIFICATION:  
; INFORMATION FOR SEQ. ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1222 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-682-517-15

Query Match 1.3%; Score 150.5; DB 2; Length 1222;  
Best Local Similarity 20.0%; Pred. No. 0.00095;  
Matches 185; Conservative 110; Mismatches 345; Indels 287; Gaps 45;

Qy 1262 DETLYRYRDEYDEGAGPSGTQHEAVPGDDNDGTSSTSSSYSDVTVNVRDVGISTNG-EVT 1320  
Db 164 DFALVFKRTVDKVEGETP-----EEAAFYKAINNTTVEFTFEETVNTVQALNFKIEGLEIK 219

Qy 1321 GEEETHSPRSVQYTYTVEEVEAPSAABROQDPPSG--SGTADAM-AFVESYKKGVDVDFH 1377  
Db 220 NASVKQTNKKVVLVLTTEAQTADKAEVVLTDGETIGGFKGVAAVVPTKVELYSSAV----- 274  
Qy 1378 QOSSGETAREVEVDGK-----GLLPESVVGAEAPTQERG--- 1410  
Db 275 ---OGKLCQEKVKQAKVTVABEQSKAGIPVTFVPGNNNDGVVP--TLTGEALTNEGIAT 330  
Qy 1411 ----RAADGNTAOTAVNEGDR-----EPVOSSLVSSPOADIPKVTQ 1447  
Db 331 YSYTRYKEGTDDEVYATAYAGDRSKFSLGVYVFWGVDTILSVEEYTTGASVNGGANKTYKVTY 390  
Qy 1448 SEVHAQK-EVKQEVPLATVSGATPIVDEKPAKPSVTVTRGVKIIDKGAHVAEKKQVOVE 1506  
Db 391 KNPKTGRPEANKTFNVGFVENMN-VTSDKVA-NATVNGVKALQLSNGTA--LDRAQIITD 446  
Qy 1507 QPKORSLTINSGKAGKQLCMFRTSCGVQOLDVYNEATATRFSNAFTVDNLKGRSAVFF 1566  
Db 447 SKGEATFTVSGTNAAVTPVY-----DLHSTNNSTSNKKYSASALQTT---ASKVTF 495  
Qy 1567 SKLGEQYTY-----NGGSHVSSGWPRALEDILTAKYPSVDFHCLVOKYKMG----- 1614  
Db 496 AALQAEYTIETRADNAGE-----VAAIGATNGREYKVIIVKDAGNLAKNEI 542  
Qy 1615 -GVFPFHADDEEC-----YPSONPILTVNLVG-KA-NFSTKCR-KGKVMVINVASGD 1662  
Db 543 VNVAFNEKDRIIVTNAKEVDTPDTAVYFTGDKAKQISVKTNDKGEATFVIGSDIVN 602  
Qy 1663 YFLMPCGQFQRLHLSVNSIDEGRISLTFRATRRVFGVRMLQLAGVSDKSPGVN--- 1719  
Db 603 DYATPIAW-----IDINTSDAQ-----GDIDEGEPRAVAPIS 635  
Qy 1720 --QOPOSOGA-----TRTITPKSGGKA-----LSESGREVKG-----RSTYSIW 1757  
Db 636 YFOAPYLDGSAIKAYKKSDLNAKAVTKFDGSETAFVAAELVNQSGKKVTGTSIKKATYTI- 694  
Qy 1758 CEQDYVRKCEWLADNPVMALEPDYTPMTFEV--KTCTSEDVAVVEYLKYLAIIGIERTYR 1815  
Db 695 ----YNTGANDIKVDNQVISPNRSYT-VTYEATLSSTGT----- 728  
Qy 1816 ALLMARNIAVTTAEG---VLKVPNOVYESLPGFHYKYSGTDLIF---HSTQDGLRVRLDP 1869  
Db 729 VITPAKMLEVTSVDGKTTAVKV-----IATGIAVNTDGKYAFTAKEATATFTATNEVP 782  
Qy 1870 YVLAEEKGIFTKGRDVAVALGDNLFVCDIILVFHDAILNLIGALKVARGCMVGESPKSF 1929  
Db 783 N---SYTCVATQFNTADS-----GSN-----SNSIWFAGKNPVKYAGVSGKTYKYF 825  
Qy 1930 -----EYKCYNAPPGGKTT-----TLVDEBFVAKSPNSTA----- 1958  
Db 826 GANGNEVFGEAAWEALLTOYATEGOKVTISYNDGDTVTFKVISAVNSSTEAIKPVAPTT 885  
Qy 1959 -----TITANVGSSSEDINMAVKRDPNL--EGLN-SATTVNSRVNIVIRGMYKRV 2006  
Db 886 PAAPTGTGALTTPAAGGLVLDLTATNTLGLSLADLNVSATVTDATVS----- 935  
Qy 2007 LVDEVHMHQGLLQGVATGASEGLF 2033  
Db 936 LKDSAN-----NSLSLTLLVETGANTGVF 958

RESULT 8  
US-08-682-517-9  
; Sequence 9, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible







QY 1572 GYTYNGSHVSGWPRALIEDILTAIKYPSVDFHCLVQKMGCGVFFRHADDECIYPSNP 1631  
DB 1002 VVEVDLEHVOHIIAGAGNDISITG---NAHDFLA-----GGG-----GDD----- 1039  
QY 1632 ILTVNLVGGKANEFTCKRGGKVMVWVASGDYFLMPCGCFQRTHLHSVNSID 1682  
DB 1040 ----RLDGGAGNDTLVGEGQNTVIGGAGDDVFLQDLGWSNQLDGGAGVD 1086

RESULT 11  
PCT-US95-10661A-6  
: Sequence 6, Application PC/TUS9510661A  
: GENERAL INFORMATION:  
: APPLICANT: Washington University, et al.  
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
: STREET: 4 Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: PCT/US95/10661A  
: FILING DATE: 16-AUG-1995  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/296,791  
: FILING DATE: 25-AUG-1994  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Trecartin, Richard F.  
: REGISTRATION NUMBER: 31,801  
: REFERENCE/DOCKET NUMBER: FP-59941/RFT  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 781-1989  
: TELEFAX: (415) 398-3249  
: TELEX: 910 277299  
: INFORMATION FOR SEQ ID NO: 6:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1848 amino acids  
: TYPE: amino acid  
: TOPOLOGY: unknown  
PCT-US95-10661A-6

Query Match 1.2%; Score 142.5; DB 4; Length 1848;  
Best Local Similarity 18.3%; Pred. No. 0.011;  
Matches 250; Conservative 151; Mismatches 435; Indels 531; Gaps 63;

QY 532 NP--VLDVKDKRRINEILFLSTAG---GDSYVSSDLLLTA-----ASKSVSY 574  
DB 91 NPQYVGVKHYNSGVSELHFGNLGNMNGNAKSHRDYSSSEENRYTYVEKNFPTEVNTS 150  
QY 575 CSRSQNCDSRADAGFM--VD--VYDISPQQVAEAMDKKGLVAFDIALMFPVELLYGNG- 629  
DB 151 PTKBQDAQKRREDYMPRLDKFVTEVAPIEASTANNKGE--YNNSDKYPAFVRLGSGT 208  
QY 630 -----EYLEELDTLVKREGDYL-----AYNVGQCCEMY----- 658  
DB 209 QFIYKKSRYQLIITEKD-----KQNLNMDVGGDNLELVGNAYTYGIAGTPYKVNHEN 264  
QY 659 -----EHS-----PSNV-----SGFFTFYSVVRTSSGNVFKLEYEGYRCG 692  
DB 265 NGLIGFGNSKEEHSDPKILSQDPLNTNVLGDSGPLFVYDREKGLWFLGSLYD-FWAG 323

QY 693 YHHLTM--CRAOKSPGTEVYTRSL-VPSFVGKSLVPIPVVAGSSVSFKTIV-----LDS 744  
DB 324 YNKSQMOENIYKHFAEKIYQYSGASLIGSTNTQWATGST---STITGGGEPLSD 380  
QY 745 FVD-----RIYSYALNTIGTFE-NRTPEYAVCAVRSOKTHVITGSRVVSVDISPD 797  
DB 381 LTDGDKDPNHGKSITLKSGSTLTLLNNHIDQGAGGLFEGDYEKGT-----SDST 431  
QY 798 W-----GLVVA-----VMAQAIKDKRAKSIRSYNFIKASEGSLAGVEKL----- 835  
DB 432 WKGAGSVADGKTVTKVHNPKYDLAKIGKGTLLVVEGKKGNEGLLVKGDTGVTILKQ 491  
QY 836 -----FFQTVG-----DCFSNAVSVY----- 851  
DB 492 ANNVQAFSQGVISGRSTLVLDNDKQVDPNSIYFGRGRGLDNGNSLTFDHRNIDG 551  
QY 852 AKAMVHDNFNVL-----ETLMSMPRAF-----IRKVP-----GSV 881  
DB 552 ARVNHNTNTNITIGESLITNPWTITSYNIEAQDDHPLRISIPYRQLYFNQDNRS 611  
QY 882 VVTICTSGASDRLEL-----RGAFDISKETFGKRLKNSRLRVFSRAIVEDS 927  
DB 612 YVTL-KKGASTRSELQNSGESNENWLYMGRTSDAAKRVNMHNHNERMNGENGYFGE 670  
QY 928 IKVMKAMKTEDCKPLPITEDSVYAFIMGNVNVHCTRAGLLGSKATVYVSSVSKG-LV 986  
DB 671 TKA-----TQNGK-LNVTFN-----GKSDQNRFLTGTCTNLNGDLNVEKGTFL 714  
QY 987 GAATKAFSGITSFSTFSGSLFYDRGLTDERLDALVTENAINSPVGILETSRVAVSK 1046  
DB 715 GRPTPHARDIAGISST-----KKDPHF-----TEN----- 739  
QY 1047 GTKEFWEVSLNDFTTFLVRNKVLIGIFVASLCAAPIANKYRRGIAANARRAGSYETL 1106  
DB 740 -----NEVVVED--DWINRN-----PKAT-----TMNVTGNASLYSGRNV 775  
QY 1107 SSLSQAGGLRLTSST-----VSGSLVVRGRFSASVTVTRATVAKQVPLALLS 1158  
DB 776 TS-----NITASNAQVHIGYKTGDTVCVRSDDTYGYTCHNSNLSEK----- 821  
QY 1159 FSTSYAISGCSMLGIWAHALPRHLMFFGCLGTLILGARASANT-----WKFGGFS 1208  
DB 822 FNPTNLRGNVN-----LTENASFTLGKANLFGTIQSIGTSQVNLKENSHTLGN 873  
QY 1209 -----NWCAPVEVVRG-----KSVS 1224  
DB 874 VNQLNLITNGHIHLNAQNDANKVTTNTLTVNSLSGNGSEYVVDFTNNKSNVVK 933  
QY 1225 SLL-----LPITLGVSLIIRG-----LLNDTIPOLAY 1251  
DB 934 GNFTLQVADKTGEPNHNELTLFDASNATRNLEVTLANGSVDRGAWKYKLRNVGRY 993  
QY 1252 VPPVEGRNVYDETLRYRDFDYDEGAGPSGTOHEAVPGDDNDGSTSSVSYDVTVNRDV 1311  
DB 994 NPEVEKRN-----QTVDTNTTTPNDIQADAPSQAQSNNEETARVE-----TP 1039  
QY 1312 GSTNGEVTGGEETHSPRSQVQYVEE-----EVAPSAVAEROGDPG-----S 1360  
DB 1040 APATESATASEQETPRAETAQAPAMEETWTANSTETAPKSDTATOTENPNSVSP 1099  
QY 1361 AMA-----FVESVKGVDDVFH-QOSSGSETAREVDGKLLPESVVGAPTQ-ERGRA 1413  
DB 1100 KVAENPPQENETVAKNEQATEPTPQNGEVAKE-----DQPTVEANTOTNEATQS 1149  
QY 1414 DNTAQTAVNEGDRPQVSSLYS--SPQADIPKVQSEVHAQKEVK----- 1457  
DB 1150 EKGTEETOTAEKTSSEPTSVTSVNOPEKTVSQSDEDKVVEKEKEKAKVETEETOKA 1209  
QY 1458 -----QEVPL-----ATVSGA----- 1468  
DB 1210 TSKEPPKOAEPAPAEVPTDTNAAEAQAQLOOTQPTTVAARETTSPNSKPAEETQ 1269  
QY 1469 -----TPIVDEKPA--PSVTTRGVKI-IDKGKAVAHVAEKKQVQVEOP 1508



Qy	1742	--BGS	GREVKGRSTYISWCEQD	VVRKCEWL	RADNPVMALEPDYTPMTF	VVVKTTSDAV	1799	
		:	:	:	:	:		
Db	1392	TDD	GANDAKAXDTLLKAGK	NLKVCRD--	GKNTFALANDLS--	VKSATVSDKL	1442	
		:	:	:	:	:		
Qy	1800	VEVLKYL	LAIGIERTYRALLMARN--	-----	IAVTTAEGVLKVPNQVYESL	PGFH	1846	
		:	:	:	:	:		
Db	1443	SLCTNG	KNYINISDTKGLNF	AKDSKTGDDANIHLNG	IASTLTDTLLN--	--SGAT	1493	
		:	:	:	:	:		
Qy	1847	VYKSG	TDLIPHSTQDGLRV	DRDLPYVLIAEK	GFTHGKDQDAV-VALG	DNLPVCDILVFH	1905	
		:	:	:	:	:		
Db	1494	TNLGG	INTDNERKXAASVKD- -----	VLNAGW	NVRGVGPASANN	OVENIDFVATY	1543	
		:	:	:	:	:		
Qy	1906	DALNL	GALKVARCGMW--	-GESFK	SFEYKCYNAPPGGK	TTTLVD- -----	EFVKSP	1954
		:	:	:	:	:		
Db	1544	DTVDF	VGDKDITSVITESK	DNKGRTEVKI- -----	GAKTSV	KDHNGKLTGKEL	KDAN	1597
		:	:	:	:	:		
Qy	1955	NSTAT	TANVGSSD- -----	-----	INMA--	VKRD	PNLEGLNSATTVNSRVNF	1997
		:	:	:	:	:		
Db	1598	NGVWT	ETTDGKDEGNGLVT	AKAVDAN	KAGWRVKTTHGANG	DNDFATVSGTNTVF	1655	
		:	:	:	:	:		

RESULT 13

```

US-08-685-467-4
: Sequence 4, Application US/08685467
: Patent No. 6060059
: GENERAL INFORMATION:
: APPLICANT: St. Gene III, Joseph W.
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/685,467
: FILING DATE: 22-JUN-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,995
: FILING DATE: 24-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-61053-2/R/T/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1912 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-685-467-4

```

```
Query Match      1.2%; Score 138.5; DB 3; Length 1912;
Best Local Similarity 18.1%; Pred. No. 0.027;
Matches 315; Conservative 208; Mismatches 598; Indels 617; Gaps 77;
QY 627 GNCEVYLEEDTLVKREGDYLAYNVGCGMEYHSFS-----NVS 666
    ||| ||| ||| | | | | | | | | |
```

Db	168	GNVHNLNGLDST-----LPDAVNTNTGVLSSSSFTPNDEKTRAAATVQDVLNAGNMNIK	220
Qy	667	GFFTFSVYRTSSGNVNFLEYEGYRCGYHHLLTMCRAQKSPGTEVTVSLVSPFVGKSLVF1	726
Db	221	G-----AKTAGNVESVDL-----	234
Qy	727	PVAGSSVSPKTVILOSDFVDRIYSYALANTIGTFENRTFEYAVGARSRQKTHVITG----	782
Db	235	-VSAYNNVEF--ITGDKNTLDVVLTAKENXKTT--EVKFTPKTSV1KEKDGKLGFTGKNN	289
Qy	783	--SRVHISKVDISPDDMMGLVAVMAQA1KD-----RAKSIRSYN-----	820
Db	290	DTNKVTSNTATNDTDEGNGLVT---AKAVIDAVKAGNWRVKTTTANGONGDFATVASGTN	346
Qy	821	-FIKASEGSLAGVFKLFQFQVGCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPK	879
Db	347	VPFESGDGTTASVTK-----DTNGNGITVYDAKVGDLKFDSD-----KKIVA	390
Qy	880	SVVVTICTSGASDRLELRGAFD1SKETFGKKLUNS-----RLR	917
Db	391	DTTALTITGTKG-----VAEIAKEDBKKLVLNAGDLVTALGNLSWKAKAEADTDGALE	442
Qy	918	VFSRAIVEDS1KYMKAMKTEDGRLP1TED-----SVYAFIMGVSN-----	959
Db	443	G1SK---DOEVKAGETVTFRAGNKLKAVKQDGANFTYSLQDALTG1TSLTGTTNGGND	499
Qy	960	-----VHCTRAGLGGSKATVSVSSVSKGLVARG--AATKAFSGITSEFSTGSLFYDR	1009
Db	500	KTVINKOGLIITPAG--NGGTTGTNT1SVTKDGIKAGNKAITVWASGLRAY-----	548
Qy	1010	GLTEDRLDALVRTENA1NSPV-----GILETSRVAVSK-----V	1044
Db	549	--DDANFDVLNNSATDLNRHVEDAYKGLLLNLNEKNANKQPLVTDSTAATVGLDKRLGW	605
Qy	1045	VA---G7KEFWSEVSLNDFTTFLVRKNKVLIGFVASLGAAPTAWKYRGIAANARRVAGS	1101
Db	606	VSTRNGTKESNOVQKQADE1LF-----TCAGAAVTVTSKEN-----GK	643
Qy	1102	SYETLSSLSSQAAGGLRG-----LT-----SSVTSVSGSLVVRGFPSSA	1139
Db	644	HTIVSVAE7KADCGLEKGDGDT1KLKVQNDQNDVLTVGNNGTAVTKGFTFVTKGATDA	703
Qy	1140	-----VTVTRAT-----VAKRQVPLALLSFSTSYATISGCSMLGIWAHALPHRLMFFFG	1187
Db	704	DRGKVTVKDATANDAKKVATVKDVATA1NSAATFVKTE-----N	743
Qy	1188	LGTL-----GARASANTWKFEGFSNNKCAVPVVRGKSVSLLLP1TLGVSL	1236
Db	744	L1T5IDEDNPTDKDALKAGDT1LFKAGKN-----LKVKRDGNK1T-----FDLAKNL	793
Qy	1237	I1R--GLLNDT1-----POLAYVPPVEGRNVYD1ETLAYRDFDYDEGAGPSG	1281
Db	794	EVKTAKVSD1LT1TGGNTPGTTATPKVNI1TADGLNPAKET-----ADASG	841
Qy	1282	TQHEAVPG-----DDNDGSTSSVSSYDV-----VTNRVDR-----G1STNGEVTGEE	1323
Db	842	SKNVYLKGIATTL1EFSAGAKSHSHVDLNVDATKSKNAASIEDVLRAGNIOGNNGNVYV	901
Qy	1324	ETH1SPRSVQ1TYVEEVAPSAVAERQDPSGSGTADAMAFVESVKVKGVDDVFHQOSSGE	1383
Db	902	ATYD--TVNFT--DDSTGTT1TVTVOKADGKGA-----DVK1GAKTSV1KDHNGK	947
Qy	1384	--TAREV-----EVDGK-----GLLPESVWGEAPTQE-----RGRAADGNTAQATVN	1423
Db	948	LFTGKDLKDANNGATVSEDDGKDTGTLVTAKTVD1AVNKGSRVTVGBGATETGATAVN	1007
Qy	1424	EGDREPQVS--SLVSSPQAD1PKVTOQSEVHAQKEVQEP1ATVSGAT1PVDKE-----PAP	1478
Db	1008	AGNAETVTS1SNFRKNGNATTATVSKDNGN1NVKYDV---NVGDGLIGDKK1IVADTT	1064
Qy	1479	SV1TRGVK11DKKAVAHVAEKQOVQE-----OPKORS1T1NEGK	1519
Db	1065	TL1TVTGKSVSPAGANSVNNKK1LVNAEGLATL1NLSWTAKADKYADGESEGETDOEVK	1124



Db	1062	-----KNVSLCTPSKPQHPALQNWNISTRPGEFTALVCP	1097
QY	1356	SGTADAMAFVESYKGVDDVFHQSSGETAREVEVDGK-----GLLPESV-	1400
Db	1098	AGKSTILSL-----QREYDPTAGSVQLDGGDIREVAVPQHRGRLGLVPEPD	1145
QY	1401	-----VGEAPTQ-----ERGRADGNTAQTAVNEG	1426
Db	1146	LPFGSISYNIIGLGAAPQGLVTRDDIEKICAKCGIHEFTMSLPEGYSTECGTNGSKLSGGQ	1205
QY	1427	RE--PVQSSLVSSPADI--PKVTQSEVHAQKEYKQEVPLATVSGATPIVDEKPA	1482
Db	1206	KQRIAVARALIRSPVLLLDYTSALDAHSEQQIKAEADVAGSDRTTIVVAHRLSTVQNA	1265
QY	1483	RGVKIIDKGVKAV	1494
Db	1266	DRIFVFDGGRV	1277

RESULT 15

```

PCT-US95-10661A-5
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/REF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
PCT-US95-10661A-5

```

Query Match 1.28; Score 135; DB 4; Length 1702;  
Best Local Similarity 19.98; Pred. NO. 0.045;  
Matches 181; Conservative 97; Mismatches 340; Indels 292; Gaps 42;

[illegible]

Qy	834	KLFQTVDGCFSNVSVYAKAMVHDNFNVLETMLSMSPRAFIKVKYPGVSVVTICTSGASDR	893
Db	586	-----NPFARIRIKDGGQLYLENTNYVALRK-----GASTR	618
Qy	894	LEL-----RGAFDISKETFGRLKLNSLRVESRAIVEDSIKYMKAMKTEDG	939
Db	619	SPLPKNGSESNWLYMGKTSDAKRNVWNHINNMKGFCNYGGE-----EEG	667
Qy	940	KPLPITEDSYAFIMGNVSNVHCTRAG-----LLGSKATVVSSVSKG-LVARGAAT	990
Db	668	K-----NNGNLNVTFPGKSEQNRFLLTGTGTLNGLDKLKEVGKGTLSGRPT	712
Qy	991	KAFSGITSFSTGSLFYDRLGTEDEBL-----DALVR-----TENA-----I	1027
Db	713	PHARDIAGISITKK-----DOHPAENNEVVVEDDWIIRNFKATNINVTNATLYSGRNVANI	769
Qy	1028	NSPVGILETSRVAV-----SKVVAGTKFEWSEVSLNDF-----	1060
Db	770	TSNITASDNKVHIHGKAGDYCVRSIDTYGYTCITDKLSOKALNSFNATVSCNVNLSG	829
Qy	1061	TTFVLNRKNVLIGIFVASIGAAPIA-----WKY-----RRGIAANARRYAG-	1100
Db	830	NANFVLGKANLFGT-ITSGTNSQVRLTENSHWLHTGDSNVQNLDKGHIHLNAONDANK	888
Qy	1101	-SSVETILSSLSSQAAGGLRCLTS-STVSOGSLVVRGFSSAVTVTRATAKRVQPLALLS	1158
Db	889	VTTYNTITVNLSLNGSGSFYYLTDLSNKQGDKVVVVTKSATGNFTLOVADKTCEPTKNETL	948
Qy	1159	FSTSVAISGCSMGIWAHALPRHLMAFFGLGLTLLGARASANTWKF-----GGFSNNWCA	1212
Db	949	FDASNATR--NNLN-----SLVGNVTDLGNKWYKLRNVNGRD---LY	987
Qy	1213	VPEVWRGKSVSSILLPIPLGVSLIIIRLLNDTIPO---LAYVPVEGRNVYDETLRYIRD	1270
Db	988	NPEVEKRNQVDT-----TNITTPNNIQADVPSVPSNN---EELARVETP	1029
Qy	1271	FDYDEGAPGSCTOHEAVPGDDNDGSTSVSSYDVVTNVNRDVCISTNGCVTGCEEETHSPRS	1330
Db	1030	VPPPAPATPETTTETAENSKEOSKRTVEKNEODATETT-----AONGEVASEAKP----	1080
Qy	1331	VQYTYVEEEVAPAANAERQDPGSGGTADAMAFVESKVGVDVFHQSSGETFAREVEV	1390
Db	1081	VKANTQTNEVAQSSETETEQTTEIKETA-----KVEK-----BEKAKVEK	1121
Qy	1391	DGKGLLPESVVGAPQERGRAADNGTAQTAVNEGDRPVOSSLLVSPQADIPKVTQOSEV	1450
Db	1122	EERAKVEKDIEQAPOM-----ASETSPKAQPAKPEVSTD-----KVEEQTV	1165
Qy	1451	HAAKENVQEIVLATVSGATPIVDEKPA----PSVTTTRGVKIIDKGKAVAHAEEKOVO---	1504
Db	1166	QAQPOT-QSTTVAAAEATSP--NSKPABETQPSKT-----NAEPVTPVVSNQNOTENT	1215
Qy	1505	VEQPKQRSLT	1514
Db	1216	TDQPTEREKT	1225

Search completed: January 31, 2001, 15:27:41  
Job time: 110 sec

**This Page Blank (uspto)**





Db 401 -----TGDASCFAIRRALSAYS----- 417  
Qy 879 GSVVVTICTSGADRLEURGAFDISKETFGKLNRLRFSRAIVEDSIKVMKAMKTED 938  
Db 418 -----NSALR-----RKIIIE----- 427  
Qy 939 GKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSYSGKGLVARGAATKAFSGITS 998  
Db 428 -----FIFGNIHP----- 436  
Qy 999 FFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVYAGTKEFHSEVSLN 1058  
Db 437 -----FD-----VAVIETNEVA-----PEPLSPEVDID 459  
Qy 1059 ---DFTTFVLNRKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSVETLSS--LSSQA 1113  
Db 460 VDCDF----- 474  
Qy 1114 AGGLRGLTSTVSGSLVVRGFSASVTVTRATVAK--RQVPLALLSFSTSAISGCSML 1171  
Db 481 RPGLUI-----RGS-----RRSSNELTSLVKVVKLARRIPRLLFRLRNFVAYF----- 523  
Qy 1172 GIWAHIALPRHLMFFEGGLTLLGARASANTWKFGGFSNNWCAPVEVWRGKSVSLLLPIT 1231  
Db 524 --VERRLASKRUKTFIGLARL-----FDNFS-----LTSVVYLLQYDSVLNAP- 565  
Qy 1232 LGVSLIIRGLLNDITIPOLAYVPPVEGRNVYDETLRYRYRDFDYDEGAGPSGTQHEAVPCDD 1291  
Db 566 IDVELI---LLN----- 574  
Qy 1292 NDGSTSSVSSYDVVTNRDVGISTNGEYTGTEETHSPRSVQYTYVEERVAPSAVAERQG 1351  
Db 575 ---SGSVNVLPLYSWR-----GSLTKLAE-----AIV----- 599  
Qy 1352 DPSSSGTADAMAFVESYKKGVDVDFHQOSSG--ETAREVEVDGKGLLPESVVGAPQER 1409  
Db 600 ---SGSFA---SFLGRMCCRVSJDMCSSSNAGCFMSPVTKGKFPVPPSSSGSTASMYER 653  
Qy 1410 GRAADGNTAQTVAVNEGDPREPQSSLSVSSPOADIPKVTQSEVHAQKEVKQEVPLATVSGAT 1469  
Db 654 LEALESDI-----REHVLSTC-----RVGSDE----- 675  
Qy 1470 PIVDEKPAPSVTTRGVKIIDKGAVAHAEKKQVQVEQPKQKRSLTINEGKAGKQKLCMFRT 1529  
Db 676 -----EERPKPE----- 682  
Qy 1530 CSCGVQLDVYNEATIRFSAFTFVDNLKGRSAVFFSKLGEQTYNGGSHVSSGWPRAL 1589  
Db 683 -----VTEPGIHTS----- 692  
Qy 1590 EDILTAKIPSPVEDHCLVKQYKMGGVFPFHADDEECYPSDNPILTVNLVGRANFSTKCRK 1649  
Db 693 EDVV-----PIRSH-----SOPLSGGEGSYEDRE-----ENE-----RAN----- 723  
Qy 1650 GGVVWVINVASGDYFLMPCGQFQTHLHVSNSIDEGRTISLTFRATRRRVFGVGRMLQLAGGV 1709  
Db 724 -----LLP-----HVSQIVSERRG--LETARNKRTLHGVSEFLNAINTS 761  
Qy 1710 SDEKSPGVNQPOSQGATRTITPKSGKALSEGSGRE---VKGRSTYSINCE--ODYVRK 1765  
Db 762 NEOPRPRIIVDHSPESTRALTNSV-----REFYVLOELALFELCKLREYDYQ 807  
Qy 1766 CEWLRAQNPVMALEPDYTPMTFEVVTKTCTSEDVAVVEYLKYLAIERTYRALLMARNTAV 1825  
Db 808 LKVANFNREQECLDKD-----EDMFV----- 829  
Qy 1826 TTABCVLKVPNQVYESLPGEHYVYSGTDLIFHSTQDGLVRVLDLPVLAIEKGIFTKGDV 1885  
Db 830 RAGGVVSGRN---SRULPKHF--KGHEFCFRS--GGL-----VPY----- 863  
Qy 1886 DAVVALGDNLEVCDDILVFDHDAINLIGALKVARCGMVGESKSFYEK-----CYNAPP 1938  
Db 864 -----DGTSRVDTI--FHTQTNFVSA--NALLSGYL--SYRTFTFTNLSANVLLYEAPP 911

Qy 1939 GGGKTTTLVDFEVK--SPNSTATITANVGSSEDI-----NMAVKKRDPNLEGLNSATTVN 1991  
Db 912 -GGKTTTLIKVFCETFSKVNLSLITANKSSREEILAKVNRIVLVDGDTPLQTRDRILIID 970  
Qy 1992 SRVNVFIVRGMYKRYL-VDEVMMHOGLLQGVFATGASEGLFFGDGINQIPIFINREKVF 2050  
Db 971 SYLMN--NRGLTKVLYLDECFMVHAGAACIEFTKCDASAILFGDSFOIRY---GRCSE 1025  
Qy 2051 MDCAV-----FVPKESVYTSKSYRCPLDVCYLLSSMTVTRGTEKCYPEKV-----S 2098  
Db 1026 LDTAVLSDLNREF--DDESRYGEVSYRCPWDVCAMLSTF-----YPKTVATTNLYSA 1076  
Qy 2099 GKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDM-----KSLKGGKGTTPVMT 2153  
Db 1077 GQS-----SMQVRETESVDDEYSSEFVYLLTMLQSEKKDLLKSFGRKRSRVSEKPT-VLT 1130  
Qy 2154 VHEAQGTFSDVFLRTRTKKADDSLFTKQPHILVLGLSRHSRLSVLAALSCKLDDKVGTYIS 2213  
Db 1131 VHEAQGETYRKVNLVTRTKFQEDDPPRSNNHITVALSRHVESLTVSVLSSKRDDAIAQAIV 1190  
Qy 2214 DA 2215  
Db 1191 KA 1192

RESULT 2  
JQ0096  
hypoetical 176K protein - papaya mosaic virus  
C:Species: papaya mosaic virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Jun-2000  
C:Accession: JQ0096  
R:Sit, T.L.; Abouhaidar, M.G.; Holy, S.  
J. Gen. Virol. 70, 2325-2331, 1989  
A:Title: Nucleotide sequence of papaya mosaic virus RNA.  
A:Reference number: JQ0096; MUID:89381685  
A:Accession: JQ0096  
A:Molecule type: genomic RNA  
A:Residues: 1-1547 <SIT>  
A:Cross-references: GB:DL13957; GB:D00580; NID:g222330; PIDN:BAA03050.1; PID:g222331  
A:Note: This protein shares two major domains of similarity with other potexvirus pr.  
C:Comment: The genome is a single-stranded, positive-sense RNA.  
C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

Query Match 3.5%; Score 402.5; DB 2; Length 1547;  
Best Local Similarity 24.0%; Pred. No. 1.2e-16;  
Matches 177; Conservative 87; Mismatches 230; Indels 243; Gaps 33;

Qy 1432 SSLVSSPOADIPKVTQSEVHAQK-EVKQEVPLATVSGATPTVDEKPAPSVTTRGVKIIDK 1490  
Db 466 SLLQKEPEVMEAKEAEATDEPQRPVEKQEAFASTSGRAEIQEDPATK-----K 515  
Qy 1491 GKAVAHAEKKQVQVEQPK-QRSLTINEGKAGKQKLCMFRTCSGVQLDVYNEATIRFS 1549  
Db 516 GK-----EENPNRDL-----LCPGLHLKIKN-----AEFP 542  
Qy 1550 N--AFTFVDNLKGRSAVFFSKLGEQTYNGGSHVSSGWPRALEDLTA--IKYP-SVFDH 1604  
Db 543 ELPVLDHPDHLTKRAKWFPSKDGKPYSTGGSHASRGWPNWLEKILAAATEIKEPLPEFNQ 602  
Qy 1605 CLVOKYKMGGVPHHADDEECYPSDNPILTVNLVCKANFSTCKRKGKVVWVINVASGDYF 1664  
Db 603 CLVQFKLQAAPFPHRDDPCYFKGHQVLTINHSCELTQIACQKG--KASITMGFGDYI 660  
Qy 1665 LMPGCFQTHLHVSNSIDEGRTISLTFRAT-----RRVF 1697  
Db 661 LSPVGFQESHKHAVSNNTTGGRVSLTFRCTVQONKFNDDGSMALDNLNLPKAWIPKLQNLG 720  
Qy 1698 GVGRLQ-----LAGGVSDKSPGVNQPOQS-----QGATRTITPKSGGKALSE 1742  
Db 721 FQGRLOQYDPNGALISPIETIEIRS--MPCKPEGVVYVYKTLIDGLARAPTYPSPNPIRAR 778

QY 1743 GSGREYKGRSTYSIMCEODYVRKCEWLADNPVMALEPDYTPMTPEVVKGTSEDAAVVEY 1802  
DB 779 AYTSVKNCRICALLRQCK-----EW-----GCRFDALVEA 810  
QY 1803 LKYLAIIGIERTYRALLMARNIAVTTAEGV-LKVPNOVYESLPGFHHYKSGTDLIFHSTOD 1861  
DB 811 GK-----RELAIIVHAGAGSGKSOALQTL----- 835  
QY 1862 GURVRLPVLIAEKGIFTKGKDVDAVALGONLFCVDDILLVFDHAINLIGALKVARCOM 1921  
DB 836 -----IKDNP-----ELDITVVLPNTN-----ELRUDWLKRLPKAPQ 866  
QY 1922 VGESKSFYKCYNAPPGGGKTTTLVDFVKSPN-----STATITANVGSSSEDIN 1971  
DB 867 -----EKPTEF-KALLAPP-----TPVIFDDYKGLPAGVEAFCLYFTVQVLIITGDC----- 916  
QY 1972 MAVKROPNLEGLSATTVNSRVNFIVRMYK--RVLVDEVHMHQGLL-QLGVFA--T 2026  
DB 917 ----KQSVHHESNENATSS-----IEPLVKEASELCRYINATHRNKKDLANKLGVYSEKT 969  
QY 2027 GASE-----GLFFGDINOIPINREKVF--MDCAVFVPKKESSVYVTSK 2068  
DB 970 GUTEVTHGTPPIGLHM-----LVPSLYKKQAFSEMCHKVSTYAGCGGITAPKIQIILLTEE 1025  
QY 2069 SYRCPLDVCYLLSSMTV 2085  
DB 1026 TSLCSREVLTYALSRAV 1042  
  
RESULT 3  
JQ1734  
genome polyprotein - shallot virus X  
N:Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: shallot virus X  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: JQ1734  
R:Kanyuka, K.V.; Vishnichenko, V.K.; Levay, K.E.; Kondrikov, D.Y.; Ryabov, E.V.; Zavriev  
J. Gen. Virol. 73, 2553-2560, 1992  
A:Title: Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal cistron closed  
A:Reference number: JQ1734; MUID:93019008  
A:Accession: JQ1734  
A:Molecule type: genomic RNA  
A:Residues: 1-1718 <X>  
A:Cross-references: GB:M97264; NID:g295078; PIDN:AAA47787.1; PID:g295079  
C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase  
C:Keywords: ATP; nucleotidyltransferase; P-loop; RNA biosynthesis; RNA replication  
F:915-922/Region: nucleotide-binding motif A (P-loop)  
F:978-983/Region: nucleotide-binding motif B  
F:921/Binding site: ATP (Lys) #status predicted  
  
Query Match 3.5%; Score 400; DB 1; Length 1718;  
Best Local Similarity 23.5%; Pred. No. 2e-16;  
Matches 187; Conservative 114; Mismatches 290; Indels 206; Gaps 37;  
  
QY 1530 CSCGVOLDVYNEATIAFRSNAF-----TFVDNLKRSVAFESKLGEGYTYNGGSHVSSG 1584  
DB 583 CBCGTEITV-----NSFGRAIEVAGVNLTDHMKGRLLAAAFYSRDQGYSTYGYSHKSGQ 635  
QY 1585 WPALEDILTAI-KYPSVFDHCLVOKYKMGGVPHADDEECYPSDNPILITVNLVGNKAF 1643  
DB 636 WLEGDLKILIEACGEKPTTYNOCLOVKYQGGSRIGFHSDEQAIYKGNKILTVNAAGSGTF 695  
QY 1644 STKCRKGGKVMVINVASGDYFLMPCGCFORTLHLSVNSIDEGRISLITRATRRVFCVGRML 1703  
DB 696 GIKCAKG--ETTLNLEDGDYFQMPSGFQETHKHNVAVTP-RLSFTFFST----- 742  
QY 1704 QLACGVSDKSPGVNPOQSGAATRTITPKSGGKALSEGSGRE-----VKGRSTYSIW 1757  
DB 743 ----VVNSQKKPAEPEKLNQNNACPKPSDP-----SNASGKHKKTHPAKAGN----- 786  
QY 1758 CEQDYVRKCEWLADNPVMALEPDYTPMTPEVVK-----TGTSDAAVVEYLKYL 1806

DB 787 -----KSSSP--NLEPLDAP--TVEILKLHGFTALTPOHDGTQOIRPVVYFNK-- 839  
QY 1807 AIGIERTYRALLM-----AR-----NIAVTTAEGVL--KVPNQ-----VYESLPGFHHYK 1850  
DB 830 DIHLRRKAVKTMSPARPFPFDLATSLRGIYTHKIDNRRATAYMSDVKNNTLGLVLPKL 889  
QY 1851 GTDLIFHSTODGL--RVRDLPYVLIATGEGIFTKGKDVDAVALGONLFCVDDILLVFDH 1907  
DB 890 DRDLL--SSWALAEITTTREVAVLAIHGAGGACKSALQELLRSSPEL--ADSIINIVVPT 945  
QY 1908 INLIGALKVARCGWGESFKSEYKCYNAPPGGGKTTTLVDFVKSPNSTATITANVGSS 1967  
DB 946 INLANDWKAKLPQMDPRRVMTFOKACERE-----CKSVTIFDDYKGLP--AGFVDAYLAIK 999  
QY 1968 EDINNAV-----KKRDPNLEGLSATTVNSRVNFIVRGMY----- 2003  
DB 1000 VVVELAILTGDRQSTHHQERESQISSLOSNTAQFSKYADYLVNATHRQPRRLANPIKVH 1059  
QY 2004 -----KRVLVDEVHMM-----HOGLL-OLGVFAT--GASEGLFFGDINOI-- 2040  
DB 1060 AEROLGGAVLKANIVPDLAMVLVPAPFRSGLITDLGRHAMTYAGCQGLTLNHLTIILDKD 1119  
QY 2041 -PFIINREKVERMDCAVFVPKKESSVYT-----SKSVRCPLDVCYLLSSMT--VRGTEKCY 2092  
DB 1120 TPLCSDVELY----TAFSRASESITFVNTSHDNPAFLAKLDATPYLKTLSISVREDEE-- 1173  
QY 2093 PEKVVSQKDPVVRSLSKR-PICTDDVAEINADVAVLCMTQLEKSDMKRSLAGKGETPV 2151  
DB 1174 ----AGADCPATEPLVKDVPKTHIPVANDKVQLESGKTEAMEDKDTRELMSGEEXTNLM 1228  
QY 2152 MTVHEAQGTFSDVVLFRTKKA--DSSLFTKQPHILVGLSRHRSLSVLAALSSKLDKVG 2210  
DB 1229 QTQDPV-----VQLFPHQQADEALF-----KI--TIGE 1255  
QY 2211 VISDASQSVSDALLHT 2227  
DB 1256 RIRMATPEONAKQLRHT 1272  
  
RESULT 4  
S77908  
hypothetical protein 1 - grapevine trichovirus B  
C:Species: grapevine trichovirus B  
C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 20-Jun-2000  
C:Accession: S77908  
R:Saldarelli, P.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: S77908  
A:Accession: S77908  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1707 <SAL>  
A:Cross-references: EMBL:X75448; NID:g1220265; PIDN:CAA53196.1; PID:g1061119  
C:Superfamily: grapevine trichovirus B hypothetical protein  
  
Query Match 3.4%; Score 391; DB 2; Length 1707;  
Best Local Similarity 20.1%; Pred. No. 7.4e-16;  
Matches 218; Conservative 137; Mismatches 306; Indels 422; Gaps 41;  
  
QY 1315 TNGEVTGEE-----THSPRSVQTYVEEVAPS-----AAVAERQDPSGSGTADAMAF 1364  
DB 338 TNGEESTAEALFVAQLSRQIQETKLYDKMNFSLKECIWSAVSSAMGDSLVVFFEDKARFY 397  
QY 1365 VESVFKGV-----DDVFHQOSSGETAREVEVDGKGLLPSSVGEAPTQER 1409  
DB 398 NESLESFILNCGPAKIEIKRVYRDVAYRNQ--GYCPHLVSFDGS---PDGVMDELYLOH- 451  
QY 1410 GRAADGNTAQTAVNEDGREPVQSSLSVSPQADIPKVTQSEVHAQKEVKQEVPLATVSGAT 1469  
DB 452 -----IRMDRSSEPEYSLTSA-----EVLESSIGCR----- 477

```
Qy 1470 PIVDEKPAVSVTTCGVKLIIDKGAHVAHVAEKKQVQVQ-----PKORSLTINEG 1518
Db 478 -----ISLIGRAMLAFHQOODVMINERVVLSPLDDFRSMRKALTDG 522
Qy 1519 KAG-----KQLCM-----FRTCSGV 1534
Db 523 ILGTTCLLEWEYDPLKRYSMGLITATELNAAIIGTTKITAKKEMAKECSERYQCSCGV 582
Qy 1535 QLDVYNEAT---IATRESNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGHPRALED 1591
Db 583 RLPRKLETPKLVTE-----EFTDALKGRRAAFYSRHSWNYSYTGANHASKGWPWVSN 637
Qy 1592 ILTAIKYPSVFDHCLVOKYKMGCGPFPHADDEECYPSDNPILITVNLVKANFSTKCRKGG 1651
Db 638 IAEKLGLEDFDHCLAQIYEEDKGINENADDEPCY-TDPEVVTNLGNANFHLKC---G 693
Qy 1652 KVMVINVASGDYFLMPCGFORTHLHSVNSIDEGRISLTFR-----ATR 1694
Db 694 SESVPLSDGDVLIMPKGQFKTHKHAVTGTSGRISLTFRNGINAPDEGSKMSEYEETH 753
Qy 1695 RVFGVGRMLQLAGVSDSEKSPGVNQPOQSOGATRTTP----- 1733
Db 754 RIGG-----KGCPRGQKESVSLQIILADHMRVDLAICTSMVFAKDPRAR 798
Qy 1734 ---KSGG-----KALSEGSGREYKGRSTY-----SIWCEQDYVRKCEWLRA 1772
Db 799 EEVRKGGMTLGTFCVLKLSNLGAYIESERGLVNGAYKELSCYAEDDHISEWSGPRKE 858
Qy 1773 ---NPVMALEPDYTPMTFFEVVKYTGTSADAVVEYLKYLAIGIERTYRALLMARNATVTA 1829
Db 859 TTFSNALSMNPDIRVWYVECEQ-----RAKKLVDSFQE 892
Qy 1830 GVLKPNQVYESLPGFHVYKSGTDLIHFSTQDGLRVDRDLPVLVIAEKGIFFKGD-VDAV 1888
Db 893 GFTGV-----CLNKFKQKSFNLV-----NGKELIDVH 921
Qy 1889 VALGDNLFVCDLILVFHDAINLIGALKVARCGMVGESFKSFYKCYNAPGGGKTTTLD 1948
Db 922 LTLG-----FAGSKSPYQC-----VL 939
Qy 1949 EFKVSPNSTATITANVGSSDINNAVKKRDPNLEGLNSATTNRSRVNFIVRGMKRVLY 2008
Db 940 KNSHYANSLVTPRKALCSDMSN---KVPDVKVWTFESAFAQO-----KKGYGLIVI 989
Qy 2009 DEVHMHQGLLQ--GVFATGASEGLFGDINOIPEFNREKVFMRDMCAVFPVKESVY-- 2064
Db 990 DEIGLLPGYIDWVHGFCYDLSL--LVLGDPLOCEYHSAKADHF-----FLGQESVFKK 1041
Qy 2065 -----YTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVPVRSLSKRPIGTDD 2118
Db 1042 FKHCNLYLKSRLPRN--OKLFEICDGAE---GEGVTENKPA----- 1081
Qy 2119 VAETINADVYLCMTOLEKSDMKRSLKGGKETPVMTVHAEAGKTFSDVVLFRKKADSL- 2177
Db 1082 -----KDI/TLCASQ-----RRKDKSDVNTVGESQGLSANRVNILLDK--DWSLV 1124
Qy 2178 -----FT---KOPH-----LVGLSRHTRSLVYAALSSKLD 2205
Db 1125 NDERVIVAFTRAREINIIIGDASLVNLLKRSKASTVLKLLKILGGERVTEGLLSLRKKLP 1184
Qy 2206 DKV 2208
Db 1185 DVI 1187
```

RESULT 5

S77910

hypothetical protein 1 - grapevine virus A

C:Species: grapevine virus A

C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 20-Jun-2000

C:Accession: S77910

R:Minatru, A.

submitted to the EMBL Data Library, June 1996

A:Reference number: S77910  
A:Accession: S77910  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1707 <MIN>  
A:Cross-references: EMBL:X75433; NID:gl405614; PIDN:CAA53182.1; PID:gl405615  
C:Superfamily: grapevine trichovirus B hypothetical protein

Query Match 3.48; Score 389; DB 2; Length 1707;  
Best Local Similarity 21.3%; Pred. No. 9.9e-16;  
Matches 181; Conservative 114; Mismatches 259; Indels 294; Gaps 35;

```
Qy 1426 DREPQVSSL---VSSPOADI-----PKVTQSEVHAQKEVQEVPLATVSGATPIVDE 1474
Db 535 DIEDIQAGLEGVITEEAEALRLPPTTKPKIT--EIHMD-----DD 573
Qy 1475 KPAPSVTTTRGVKIIIDKGAHVAHVAEKKQVQVQPKQRSLTINEGKAGKOLCMFRTCSGV 1534
Db 574 TPG---TSGESDVEKFSV-----RSLC----- 593
Qy 1535 QLDVYNEATIATRESNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGHPRALEDILT 1594
Db 594 REEYSE-----KLKREVAFYSRHSKEYKNGSGSHRSLGDEALNELTQ 638
Qy 1595 AIKYPSPFDHCLVOKYKMGCGPFPHADDEECYPSDNPILITVNLVKANFSTKCRKGGKVM 1654
Db 639 ELGLDDSDHCLIQRYTAGSIGFHADDEPCYLPGGSVVTNLHGDAATFEVKENQSGKIE 698
Qy 1655 VINVASGDYFLMPCGFORTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGVSDSEKS 1714
Db 699 KKLHDGDDVYVNGPCMQQTHKRVTSHTDGRCSITLR-----NKTVDYEARKGDEDS 750
Qy 1715 PGVFNQPOQSOGATRTTPKSGKALSEGSGREYKGRSTYSIWCEQDYVRK----- 1765
Db 751 EYEDKAELEDEGI-----DYLQKNQGNMCSLK 777
Qy 1766 --CWLRADNP-VMALEPDYTPMTF-EVVKGTGSDAVVEYLKYLAIGIERTYRALLMAK 1821
Db 778 AFADHMQLSTSVTAIVNGASQPOTLREIEDGYSLATLVNLSKALDF----- 824
Qy 1822 NIAVTTABGVILKVP---NQVYESLPGFHVYKSGTDLIHFSTQDGLRVDRDLPVLVIAEKGI 1878
Db 825 PIATHGERGVAETPGSYRRLHLKITSGHVE-----PFEQVTSKGG 864
Qy 1879 FTKDKVDVAVVALGDNLFVCDLILVFHDAINLIGALKV--ARCGMVGESPKSFYKCYNA 1936
Db 865 FRE-----AMLLGDGVGV-----GHFRVDKAKADRLAQSF-----YN- 896
Qy 1937 PPGGKTTTLVDEF-----VKSPNSTATITANVGS-----SEDINNAV 1974
Db 897 ---GNTGVLLGKYNKGMKMTGEIEEPKEVLTAFGFAGSGKSHWCOTILKHCSVEKVLVI 952
Qy 1975 KKRDPNLEGLNSATTNRSRVNFIVRGM---YKRVLDEVHMMHQGLLQGVFATGASE 2030
Db 953 SPRVLRDWMVAKISKRRVVTVEVAEMDDYGCCKDIVIDEIGLLPPGVIDLVAIAHQPT 1012
Qy 2031 GLFGDINOIIFINREKVFMRDCA---VFVPKESVVTYSKSYRCPLDVCYLLSSMTVRG 2087
Db 1013 LVLLGDPLOQSYHSKRDNVVLEASQEDVFNVRKGLPYLCYSHRLPRN-CKLF-EIECMG 1070
Qy 2088 TEKCYPEKVVSG---KDKPVVRSLSKRPIGTDDVAEINADVYLCMTOLEKSDMKRSLK 2143
Db 1071 AES--EKRVYRSNRLKDEPTI-----CATRANKEE-----K 1100
Qy 2144 GKGETPVMTVHAEAGKTFSDVILFR-----TKKADDSLFTKOPHILVGLSRHTRSL---V 2196
Db 1101 GSG---WYTVSETQGLSFKSLCLIVLDEHWAKEDED-----VMVALTRSGRGEIGHV 1149
Qy 2197 YAALSSKL 2204
Db 1150 TPALKKKL 1157
```

```

RESULT      6
S01865
genome polyprotein - tobacco rattle virus (strain SYM)
N:Alternate names: RNA replicase
N:Contains: 134K protein; 194K protein; RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: tobacco rattle virus, TRV
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 17-Mar-2000
C:Accession: S01865
R:Hamilton, W.D.O.; Boccara, M.; Robinson, D.J.; Baulcombe, D.C.
J. Gen. Virol. 68, 2563-2575, 1987
A:Title: The complete nucleotide sequence of tobacco rattle virus RNA-1.
A:Reference number: S01865; MUID:88034943
A:Accession: S01865
A:Molecule type: genomic RNA
A:Residues: 1-1707 <HAM>
A:Cross-references: EMBL:X06172
C:Superfamily: cucumber mosaic virus RNA 1 protein
C:Keywords: nucleotidyltransferase
F:1-1707/Product: 194K protein #status predicted <MAT2>
F:1-1187/Product: 134K protein #status predicted <MAT1>
F:1188/Region: opal stop codon.

Query Match          2.9%; Score 338.5; DB 2; Length 1707;
Best Local Similarity 18.0%; Pred. No. 1.6e-12;
Matches 334; Conservative 209; Mismatches 514; Indels 801; Gaps 72;

QY 445 DAMIREKPS-----HRCDFVLKPREREKLRLFPPELSIQFSDSVRSSHPPFANMRSCFN 498
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 50 DDMIRERTKDPVIVHE---VLSQEQNKLMEIYPEEFNVFKDDKNMVHGFAAAERKLQA 106

QY 499 GIFSRRCGNVC-FPDIGGSFTYHYKAGHVCHNCVPDLVDKDKRRINELFLSTAGGDS 557
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 107 LLLLDRLPALQEVDDIGQWFSWTRGSKRIHSCCPNLDIRDQREISRIQLTAIGDQA 166

QY 558 YVSSDLLTE-----AASKSVSYCSRESQCNSRADAGF-----MV 592
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 167 RSGKRQMSSENLWMYDQRFKKNIAPNAVRCNTTYOGCTCR---GFSDGKKKAQAIAILH 223

QY 593 DVYDISPOOVAEAMDKKGALVPDIAMLF-PVELLYXNGCEVYLEELDTLVKREGDYLAYNW 651
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 224 SLYDFKLKDLMATVWEKTKVVVHAAMLPAESMLVDEG-----PLPSVDGYVMKKN- 274

QY 652 QCQGEWYEHSNFNSVGFF---TFSVIVRTSSGNVFKELEGYCGYHHLTMCRAQASPGTE 708
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 275 ---GKIY---FGFEKDPSPFSIHQWE-----EYKKYLL-----GKP 304

QY 709 VTYRSLVPSPFGKSLVFTP-VVAGSSVSFKTIVLSDSFVDRIYSYALNTIGTFENRTFEY 767
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 305 VSQY-----GNFYFEPWQVGRDTMLFS-----LYRTA----- 332

QY 768 AVGAVRSQKTHVTGSRVVHSHKVDISPDDMGMLVVAVMAQAIDRAKSIRSINFYKASEG 827
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 333 --GVPR-----RSLSSQEQYYR----- 346

QY 828 SLACVEKLFQTQGDGCFSNASVYAKAMVHDNFVNLETLMSPRAFIRKVPQSVVVITCT 887
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 347 -----RVIYSRWENNVAVPI-- 361

QY 888 SGASDRLELRGAFDISKETFGKRLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKLPITED 947
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 362 -----FDLVEST-----RELVKDLFVEKQFM-----D 384

QY 948 SVYAFIMG-----NVSNNWHCTRAGLGGSKATVVVSVSKGLVARGAATKAFSGITSFPS 1001
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 385 KCLDYIARLSDDQUTISNV-----KSYLSSNNWWLVFINGAAYK----- 422

QY 1002 TGSIFYORGLTEDERLDALVRT---ENAINSPVGILETISRVAVSKVAVACTKEFWSEVSLN 1058
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 423 -----NQQSVDSDQLLAOTLLVKEQVARPV-----MRELREAILTETKPI---TSLT 468

QY 1059 DFTFTVLNRNKVLIGIFVASLGAAPIAWKYRRGIIANARRYAGSSYETISSLSSQAAGGLR 1118

```



Qy	2042	FINREKVF	RMCDAVF	VPKKES	VVVYTSKSYR	CPDLVCVYLL	SSMTVRGTEK	CYPERK	VSGKD	2101
		:	:	:	:	:	:	:	:	
Db	942	FVSNP	SFVRHH	KLCK	VERKLIT	---WRSPAD	ATYCL---	-----EKYFY	KNK	985
		:	:	:	:	:	:	:	:	
Qy	2102	KP-----	VVRSLS	KKPIG	TGTDVA	EINADV	LYLCTOLEK	SDMKRS	LKGK	2155
		:	:	:	:	:	:	:	:	
Db	986	KPVK	NSRVL	RSIEV	PINS	PSV	ERNALYL	CHTQEK	AVLKA	1044
		:	:	:	:	:	:	:	:	
Qy	2156	EAGK	TFSDV	VLRTK	KADDS	LFTK	QPHI-----	LVGLSR	HRTSL	2209
		:	:	:	:	:	:	:	:	
Db	1045	EAGK	TFD	PNVY	FCRL	TRTST	LSATGR	DPI	NGVPC	1104
		:	:	:	:	:	:	:	:	
Qy	2210	TYIS	DAS	PQSV	SDALL	-----	HTF	2228		
		:	:	:	:	:	:	:	:	
Db	1105	NAC	EDAG	-NT	DD	SIL	ARS	YNHNF	1126	
		:	:	:	:	:	:	:	:	

RESULT 8  
 WMTM8T  
 180K protein - tomato mosaic virus (strain L)  
 N:Contains: 130K protein  
 C:Species: tomato mosaic virus  
 C:Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 30-Sep-1993  
 C:Accession: A04195  
 R:Ohno, T.; Aoyagi, M.; Yamanashi, Y.; Saito, H.; Ikawa, S.; Meshi, T.; Okada  
 J. Biochem. 96, 1915-1923, 1984  
 A:Title: Nucleotide sequence of the tobacco mosaic virus (tomato strain) gene  
 A:Reference number: A91984; MUID: 85157522  
 A:Accession: A04195  
 A:Molecule type: genomic RNA  
 A:Residues: 1-1615 <OHN>  
 A:Note: readthrough of the terminator UAG between codons for Gln-1116 and Gln-  
 C:Superfamily: cucumber mosaic virus RNA 1 protein

Query Match	2.7%; Score 309; DB 1; Length 1615;
Best Local Similarity	16.8%; Pred. No. 1.1e-10;
Matches	315; Conservative 205; Mismatches 519; Indels 840; Gaps

  

Qy	403	ALNQEEVYSSVVSSITNRLVLRODSALLSHDLKCLDMESORDAMIREKPSHRCDVFLKP	462
Db		: : :                         :	
Qy	2	AYTOTATSGALLETVRGNTLVNDLAKRRLDYDAV--DEFNARD---RRPKNFYSKVWSE	56
Db		:                                 :	
Qy	463	REREKLRELPESLIQFSDSVRSHPHFANAMRSC-PNGIFSRRCGNVCFFDIGSFTYHV	521
Db		:                                 :	
Qy	57	EQTLLIATKAPPEQIIFYNTQNNAVHSLAGGLRSLERLEYLMQIPIYCSLYTDIGGNFASHL	116
Db		:                                 :	
Qy	522	KAGHVNCVGNPVLVDVKDKARRINE----ILFLS-TAGGDSVYS-----SDLLTE	566
Db		:                                 :	
Qy	117	FKGRAYVHCMPNLVDRIIMRHGGOKDSIELYLSRLGRKNKHVPNQEKAFORYAEMPNE	176
Db		:                                 :	
Qy	567	AASKSVSYCRESQNC-DSRADACGFWDVYDTSPOQVAEMKKGALVEDIALMFPVELL	625
Db		:                                 :	
Qy	177	VVCHDTFQTCRHSEQCYTGVRVYAIAHSIIDYPADFEAGALLRNKHVHYCAAFHESENLL	236
Db		:                                 :	
Qy	626	YGNGEVYLELDLTLVKREGDYLAYNVGQCCEM-YEHSFSNVSGFFTFSVVRTSSGVNFKL	684
Db		:                                 :	
Qy	237	LEDSHVNLDENACFDGDRFLTFSASSTLNYSHSYSNILKYCKTYFPASNREVYMK	296
Db		:                                 :	
Qy	685	EYEGRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLFIPVAVAGSSVSFKTIVLDS	744
Db		:                                 :	
Qy	297	EF-----	298
Qy	745	FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSOKTHVITGSRVWHISKVDISPPDMWGLVV	802
Db		:                                 :	
Qy	299	LVTRVNVTWCCKFRIDTF-----LLYKGVAHKGDV-----	328
Db		:                                 :	
Qy	803	AVMAQAIKDKRAKSIRSVNYETIKASEGSILAGVKELFFOTVCGDFSNVSVYAKAMVHDNENV	862
Db		:                                 :	
Qy	329	-----SEQFYKAMEDAWH-----YKTLTAMCNSEIRLLIEDSSSVNYWPFK	368
Db		:                                 :	
Qy	863	LETLMSPRAFIRKVPGSVVVVTTCTSGASDRLELRGAFDISKETGRKLKNSRLRVFSRA	922
Db		:                                 :	

[illegible]



```
Db 835 -----PGCGK-----TKELSRVNFEDLILVPGQAAEMIRRR 868
Qy 1979 PNLEGLNSATTNVRV--NFIV-----RGMKRVLVDEVHMHGGLQLQLGVFATGASEG 2031
Db 869 ANASSIIVATKDNVTVDSFLMNYGKGARCOFKRLFIDEGMLMLHTGCVNFLVMSLCDIA 928
Qy 2032 LFFGDINOIPINREKVKFMPDCAVFPVKKESVYVTSKSYRCPLDVCYLLSS-----MTV 2085
Db 929 YVYGDTQIPYINRVGFPYPAHFAKLEVDEVRRTTLRCPADVTHFLNQRYEGHVMTCT 988
Qy 2086 RGTEKCYPEKVVSG--RDKPVRSLSRPIGTTDDVAEINADVILCMTOLEKSDMKSLK 2143
Db 989 SSEKKSVMQEMVSGAASINPVSFKLKGK-----ILTFTQ-----SD-KEALL 1029
Qy 2144 GKGRETPVMTVHEAQGTFSVDWLFRTKKADDSLFTK--OPHILVGLSRHTRSLVAAA--- 2199
Db 1030 SRG--YADVHTVHEVQGETYADVSLVRLTPTVPSIIARDSPHVLVSLSRHTRKSLKYTVVM 1088
Qy 2200 --LSSKLDD--KVGTYSID 2214
Db 1089 DPLVSIIRDLERVSSYLLD 1107

RESULT 9
S03701
141K protein - pea early browning virus
C:Species: pea early browning virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S03701
R:Macfarlane, S.A.; Taylor, S.C.; King, D.I.; Hughes, G.; Davies, J.W.
Nucleic Acids Res. 17, 2245-2260, 1989
A:Title: Pea early browning virus RNA1 encodes four polypeptides including a putative 21
A:Reference number: S03701; MUID:89202038
A:Accession: S03701
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1766 <NAC>
A:Cross-references: EMBL:X14006; NID:g61019; PIDN:CAB37343.1; PID:g4456716
C:Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.6%; Score 304; DB 2: Length 1766;
Best local Similarity 17.2%; Pred. No. 2.6e-10;
Matches 331; Conservative 237; Mismatches 551; Indels 802; Gaps 75;

Qy 400 VANALNOEEVYSSVSSITNRLVLRDQSALLSHLDTKLCDFMSORDAMIREKPSHRCDV- 458
Db 7 ISQLNIDFEGAERGQGFLDUMITKPSQLGAMMQRILTDKI---DDSLREKTRTETVVI 63
Qy 459 --FLKPREREKRLFLPELSIQFSDSVSRSSHPPFANMRSCFNGIFSRRCGNV-CFFDIDG 515
Db 64 HELLSQDKNKLMEIYEFNIVFKDKNVHGFAAAERKLAQLLMARVPKLEPVDDIGG 123
Qy 516 SFTYHVAGHNVCHVNPVLDVKDRRINILFL-----STAGGDSYVSSDL--LPEAA 568
Db 124 QWSFWLSKDRVHSSCPLIDMRKQRELQONFLRVFRDNATTSKRISDDQFDMYNAF 183
Qy 569 SKSVSY-----CSRESQNCSDRADAGEWQ-----VYDISPOOVAEMDKK 609
Db 184 KGDLDVANFVRCNNTFODCNCR---GRADGTRIGATNATLHSLYDFKLDVADAMIEK 240
Qy 610 GALVFIALMPPEVL-----LYGNGEVYLEELDTLVKRGDYLAYNVGCGGMYEHSF 662
Db 241 GTKELHAAMLFAPEAIEKEGPLPSVDGYERKEGSLISSEKIEFFGN-NDPSYAIHDW 299
Qy 663 SNVSGFTTFSVVR-----TSSGNVPKLEYGYRCYHHLTMCRAQKSPGTEVT-----YRSL 714
Db 300 SEYK-----KYLGRGEPSSRRGHFMFPWQARGDTEFTLYRMTGVPMTNLLNGEYVRL 354
Qy 715 -VPSEVGSKLVFIPVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVR 773
Db 355 HISRWEG--WVVVPVFEIDEIT-KKLTKSSWIYEKATM-----DKCLDYV----- 396
```

```
Qy 774 SOKTHVITGSRVHVKVDISPDMMGLVAVVMAQAIDRAKS--IRSYNFIKASEGSLAGV 832
Db 397 -----SRLSQQLTI-----NNVKSEKSNWV----- 419
Qy 833 FKLFQTVGDCFSNAVSAYAKAMVHDENVLETLMSPRAFIKRVPGSVVVTICTSGASD 892
Db 420 -----LFINGAAVKNQSVQD-----PR----- 436
Qy 893 RLELRGAFDISKETFGKRLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAF 952
Db 437 RLQLLAQTLLVKE-----KLMRPLMTEMREKMWL----- 465
Qy 953 IMGVSNVHCTRAGLGGSKATVYSSVSKGL-VARGAATKAFSGITSEFFSGSLFYDRGL 1011
Db 466 -----RATAVNSVTGTLKVANG-----NVFD-GSL----- 489
Qy 1012 TEDERLDAALVTEINAINSPVGILETSRVAVSKVAGTKFVSEVSLNDFTTFVLRNKVL 1071
Db 490 ---RRKCLLKLQVLGTETGL-----EVLVKDAPKYIEINDYLTAI----- 528
Qy 1072 GIFVASLGAAPIANKYRRIANARRYAGSSYETILSSLSQAAGLRLCTLSSTVSGGSLV 1131
Db 529 -----FNETTEKTDGSLPDLDEA----- 546
Qy 1132 VRRGFSSAVTVTRATVAKROVPLALLSFSTSYAIS-----GCSMLG 1172
Db 547 --KRNSDKISKEAAEAQVCKMQFPKFESSDSLKEPLIRKIDGDKVKKKGQRNCGLLT 604
Qy 1173 IWAHALPHLMFFFGGLTLLGARASANTWKFGFSNNCAVPEVWVRG-----KSVSS 1225
Db 605 GWTKLL-----NNTHA--HVGWQRLFLKLSIESESE 633
Qy 1226 LLLPITLGVSLIIRGLNDTIPQLAYVPPVBSRVNYDETLYRYDFYDEGAGSGTQHE 1285
Db 634 GRKPWT---DEEIEALDDIM-----ELNDVNLEAYK----- 662
Qy 1286 AVPGDDNDGSTSSVS-SYDVVT---NVRDVGISTNGEVTGEEETHSPRSVQYTVVEEV 1340
Db 663 -----TRTVNKEFDIFTWLASTYDGLDSEKELI-----TNL 695
Qy 1341 APSAAVAERQGDPSGSGTADAMAFVESVKKGVDVDFHQQSGSETAREVEYDGKGLLPESV 1400
Db 696 LATAAVRNKK-----ALSDKLAMLIDVDVDSVNSFLRSLSDTD----- 732
Qy 1401 VGEAPTQERGAADGNT--AQTAVNEGDRPEVQSLSYSSPOADIPKYTQSEVHAQKEVKQ 1458
Db 733 -----DDSDVADCSASVSDTSCVSSVWFRP--TVPPGFE-DVNLRK----- 772
Qy 1459 EVPLATVSGATPIVDEKPAPSVTTTRGVKIIDKGAHAHAEKQVQVBPQKORSITINEG 1518
Db 773 -----GKTIVVDNDAESSSS-----SERNR----- 793
Qy 1519 KAGKOLCMFRFCSCGVQLDVYNEATIAFRSNAFTFVDNLKRSVAFSKLGEITYNGG 1578
Db 794 -----NHFAN-FEVIENT-----CRFGDAPKETGD 816
Qy 1579 SHVSGWPRALEDLITALKYPSVDFHCLV-----QKYKMGSGVPPHADDEECYP 1627
Db 817 PSVDSR-----LEFTHYL-----RCLICAQNNELGKYRDYEMG----- 850
Qy 1628 SDNPILTIVNLVGNKANSFTKCRKGGK-----VMVINVASGDYFLMP--CGFORTHLSVNS 1680
Db 851 -----VVRPGKGYPDGLGVFDLALKWIIKPPSCSYNKAFFPDVSA 892
Qy 1681 IDEGRISLTFRATRRVGVGRMLQLAGVSEKSPGVNPQOPOSGOATRTTTPKSGGKAL 1740
Db 893 KEQKG-----WIGYLVDAWGC-----RQGIDAF 915
Qy 1741 SEGSGREKVGKSTYSINCEQDYVRKCEWLNRADNPVMALEPDYTPMTVEVKTGTSEDVAV 1800
Db 916 S-----C-----ITNVANKA----- 925
Qy 1801 EYLKYLAIIGTIERTYRALLMARNIAVTTAEGVLKVPNOVYESLPGFHVYKSGTDLIHFSTQ 1860
```



```
Db 926 -----DIAIVCSQ----- 933
Qy 1861 DGLVRDLPVLIABKIGFTKGDVAVVAGDNLNLFVCDLILVHDAINLIGALKVA--R 1918
Db 934 -----TFLCNERIILK----- 955
Qy 1919 CGMVGSEFKSEYKCYNAPPGGKTTTLVDFVKSPNSTATITANVGSSEDI--NMAYKK 1976
Db 956 C-----KFKLVGPGCGKSTMIVN--TANPFDVTLSTCKEATEDLLERFAAKK 1003
Qy 1977 RDPNLEGLNSATVNSRVNFTVRGMYKRVL--VDEVHMHGOLLQLGVFATCAGSEGLFEG 2035
Db 1004 FGINKL--KRVKTVDSFLMHCSGDCVGLDHLHFDALMAHAGWVFCAQIAKAKKVICOG 1061
Qy 2036 DINQIPFINREKVRMDCAVFVKESVYVTSKSRCPDLVLCYLLSSMTVRGTEKCYPEK 2095
Db 1062 DOKQIAYKPRVSQTLRFTSLIGREDEVEEKRMYSRCPVDVALT-----DRFYTGK 1113
Qy 2096 VVSGKDKPVRSLSRPICTTDDVAEINADVYLCHTQLEKSDMKSLKSGKGTPTVMVTH 2155
Db 1114 VVT--KNSVLRSMDVKRIGSKBQVEMEHGQIYTLTQSEKDKDIANLLCQKVKSFVNTVH 1171
Qy 2156 EAQGTFSVDVLFRTKKADDSFTKOPHILVGLSRHSLVYAAALSSKLDDKVGTYISDA 2215
Db 1172 EAQGTETKRVLRVLPKPTDDVLARQEQIYVALSRHTOSLYETVK---DDEVSALIRDS 1228
Qy 2216 S 2216
Db 1229 A 1229

RESULT 10
WMTMPV
186K protein - pepper mild mottle virus (strain Spain)
N:Contains: 126K protein
C:Species: pepper mild mottle virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994
C:Accession: JQ1312
J. Gen. Virol. 72, 2875-2884, 1991
A:Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistance
A:Reference number: JQ1312; MUID:92113528
A:Accession: JQ1312
A:Molecule type: genomic RNA
A:Residues: 1-1611 <ALO>
A:Cross-references: GB:M81413
A:Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln and
C:Comment: This protein may have RNA polymerase activity.
C:Superfamily: cucumber mosaic virus RNA 1 protein
F:1-1117/Product: 126K protein #status predicted <PRO>
```

```
Query Match 2.6%; Score 297; DB 1; Length 1611;
Best Local Similarity 22.8%; Pred. No. 6.1e-10;
Matches 195; Conservative 114; Mismatches 301; Indels 244; Gaps 40;

Qy 1508 PKQRSLLTI-----NEGK--ACKQLCMFRTCSGVLQDYNPATIATFRSNAFTFVD 1556
Db 366 PKMKDMVIVPLFDVSLQNEGKRLARKEVMWSKDFVYTVLNLHRTYQSALTYANVLSFEV 425
Qy 1557 NLKCR-----SAVFF--SKLG-----EGYTYNGGSHVSSG 1584
Db 426 SIRSRIIVGTARSEWDVKALLQSLSMFTFLQTKLAMLKDDLVLVQRFVHKSLETVV 485
Qy 1585 W-----PRALEDILT-----AIKPSV---FDHCLVQKVMKGGV-- 1617
Db 486 WDEITAFHNCFPPIKERLIINKLITVSEKALEIKVPDLYVTFDRDLRVEKVEKSSVEMPL 545
Qy 1618 -FHADDEECPSDNPILTVNLVGRAN-----FSTKCRKGG-----KVMVINVASGDY 1663
Db 546 DVKKSLEAEVYNALSEISILKDSKDFDVFDSRMCNTLGVDPILVAAKVMVAVVS----- 601
```

```
Qy 1664 FLMPGCFQTHLSVNSIDEGRISLITFRATRRVFGVGRMLQLAGGVSDKSPGVPNQPOQ 1723
Db 602 -----NESGLTTLTTFERPTEA--NVALALQ-----PTITK 629
Qy 1724 SOGATRTITPCKGKALSGREVKRGSTYST--WCEODYVRKCEWLADNPVMA 1777
Db 630 EESSLIKIVSSDVG-----ESSIKVVRKSEISMGLTGNTVDSDEFORSTE-----IES 677
Qy 1778 LEPTDPTMTFEVVKGT-----SEDVVVEYLKYLAIGIERTY--RALLMARNIATVTAEGVLK 1833
Db 678 LQ-----QFHMVSTETIIRKQMHAMVYTPGLKVQCKNVLDSLVASLSAAVSNLKKIIK 731
Qy 1834 VFNQVYESLPGFHVVKSGTDLIFHSTQDGLRVLDLPYVLIABEKGI---PTKDKDVAVVA 1890
Db 732 -----DTAAIDLETKEFGV---YDVCLKKWLKPLSKGHANGVVM-- 769
Qy 1891 LGDNLFCDDIILVHDAINLIGALKVARGC---MVGESFKSEYKCY----- 1934
Db 770 --DSDYKCFVALLTYDGENIV-----CCETWRRVAVSSESLVYSDMGKIRAIRSVLKD 820
Qy 1935 NAPPGGGKTTTLVDFVKSPNSTATITANVGSESDINMAVKK-----RDPNLEGLNSA 1987
Db 821 GEPHISSAKVTLVDG--VPGCGKTKETLSRVNFEDELVLVPGKQAAEMIRRRANSSGLIVA 879
Qy 1988 TTVNSRVV--NFIV---RG---MYKRVLVDEVHMHQGLLQLGVFATCAGSEGLFGDINOQIP 2041
Db 880 TKENYRTVDSFLMNTYGRGPCQYKRLFLDRLHGLHFCVNFVGLMSLCSEAFYVGDQQIP 939
Qy 2042 FINREKVRMDCAVFVKESVYVTSKSRCPDLVLCYLLSS-----MTVRGTEKCYPEK 2095
Db 940 YINRVATFPYPKHLSOLEVDVAVETRRRTTLCRPCADITFFLNQYEGQVMTSSTVRSVSHE 999
Qy 2096 VVSGKDKPVRSLSRPICTTDDVAEINADVYLCHTQLEKSDMKSLKSGKGTPTVMVTH 2155
Db 1000 VIQG--AAVMNPVSKPLGKV-----ITFTQSDKS---LLLSRGYE--DVHTVH 1041
Qy 2156 EAQGTFSVDVLFRTKKADDSFTKQ--PHILVGLSRHSLVY-----AALSCKLDDK-- 2207
Db 1042 EQVGTEDFEDVSLVRLTPTVPVGIISQSPHLLSLSRHTSIKRYTVVLDVAVSVLRDLEC 1101
Qy 2208 VGTIISDASPSQSVS 2221
Db 1102 VSSYLLDMYKVDVS 1115

RESULT 11
WMTMS2
186K protein - cucumber green mottle mosaic virus (strain SH)
C:Species: cucumber green mottle mosaic virus
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Apr-1994
C:Accession: JQ1157
J. Gen. Virol. 72, 1487-1495, 1991
A:Title: The complete nucleotide sequence of cucumber green mottle mosaic virus (SH)
A:Reference number: JQ1157; MUID:91311400
A:Accession: JQ1157
A:Molecule type: genomic RNA
A:Residues: 1-1646 <UGA>
A:Cross-references: GB:D12505
C:Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.6%; Score 294.5; DB 1; Length 1646;
Best Local Similarity 17.0%; Pred. No. 9.1e-10;
Matches 327; Conservative 192; Mismatches 515; Indels 885; Gaps 70;

Qy 405 NOEEVYSS-----VVSSTINRLVLRDQSALLSHLDTKLDMFSQORDAMIREKSHRCDFVL 460
Db 10 NQORDAAASGRNNLVSQLASKRYDEAVRSLDHQD-----RRPKMNFSRV 54
Qy 461 KPREKRELRELPPELSIOFSDSVRSRSHPPANARSCNCFISRRC--GNVCFDFDGGST 518
Db 55 STEHRLVTDAYPEFSISFTATKNSVHSLAGGLRLLELEYMMQVPYGSPC-YDIGGNYT 113
```

Qy	519	YHVKAGHVNCVNPVLVDKVKRRI--NEILF-----LSTAGGSYVSSD	562
Db	114	QHLFGKRSYVHCNCPCLDLKDYARNVMYNDMTIQHVHRKSGCGRPLTFQIDAFRYD	173
Qy	563	LLTEAASKSVSYCSRSONCDSRADG-----FNVDDYDLSPOOVAEMDKKCALYFDI	616
Db	174	-----SSPCAVTSCDFQSCSYDPSGGRDNHAVLSHSIDIPYSSIGPALHURKNVCYA	228
Qy	617	ALMFPVLLYNGEVEYLEELDLVLRREGDYLAYNVGQCGEM-YEHSFNSVSGFFTSYVR	675
Db	229	AFHF\$KALLGSPVGNLSIGAQFRVGDGDVHFLFSEESTLHYTHSLENIKLIVMRTYFP	288
Qy	676	TSSGNV\$KLEYEGYRCGYHHLTMCR-----OKSPCTEVTYRS-----LVPSFVGKSL-	723
Db	289	ADDRFVYIKEFMVKRYDTFFFLVRADTHMLHKSYGHY\$KWKSEYFALNTPPIQDKATF	348
Qy	724	-----VFIPV-----VAGSSVSFKTIVLDSDFVDRIYSALNTIGFENR--TFEY	767
Db	349	SVWFPEAKKVLIPKFELSRFLSNWKISRMULDADFVHTI-----INHISYDNKALVKN	404
Qy	768	AVGAVRSQKTHVITGSRVH\$KVDISPDMMGLVYVAVMAQAIKRAKSIR--SYNFIKAS	825
Db	405	VQSPVESIR\$RVLVNGSV\$KSEMNVPDQLTDIS\$IF-PLVKVRKQVQIELMSDKVIVEA	463
Qy	826	EGSL---AGVFKLFFQTVGDCFSNAVSY\$KAMVHDNFNVL\$TMSMPRAFIRKVP\$VV	882
Db	464	RGLLRRFADSL\$K\$AVEGLDC-----VY-----	486
Qy	883	VTICTGASDRLELRC\$AFDISKETTCRKLKNSRFLVSRAIVED\$IKVMKAMKTEDGKPL	942
Db	487	-----DALVQTCWFDTS-----SDELKVL-----LPE	508
Qy	943	PITESDYAPIMGN\$NVNHCTRAGLGG\$KATV\$SV\$KGLVARGAATKAFSGITSPFST	1002
Db	509	PFMTF\$DYL-----EGMYEADAKTERESVSELLASGDDL\$FKKIDEIRNYS-	554
Qy	1003	GSLEYDRCLTEDERLDALVRTENALNSPVGILETSRVAV\$KVVAGT\$KEFSEVSLNDFTT	1062
Db	555	-----GV\$EFD-----V\$KFOEFC\$KELNV-----	572
Qy	1063	FVLNRKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLS\$QAA\$GLRGLTS	1122
Db	573	-----NPMLIC-----HVLEATFSQKAG-----	590
Qy	1123	STVSG\$LVVRRG\$SAVTVTRATVAKRQVPLALL\$FST\$YAISGC\$MLGIAW\$HALPRHL	1182
Db	591	-----VTVT-----	594
Qy	1183	MPFEGCLTLGCASANTWKFGG\$NNHCAPEV\$VWRCK\$SVSSLLLPITLGVSLIIRGLL	1242
Db	595	-----GLGTL-----SP\$M-----GASVA-----L	609
Qy	1243	NDTIPQLAYVPPVEGRNVYDETLYRYDFDEGAGP\$CTQHEAVPGDDNDG\$STSSV\$Y	1302
Db	610	SST-----	620
Qy	1303	DVVTNVRDVGISTNGEVTG\$EETH\$SPRSVOYTYVVEEVAP\$AAV\$AERQGP\$SGGTADAM	1362
Db	621	DVTEDM\$DI-----VLMADK\$HS-----YMSPEMARWA-----	648
Qy	1363	AFVESYKKGVD\$VFHQ\$SGSETAREVEVDGKLLP\$SVVGEAP\$TQERGAADGN\$TAQTAV	1422
Db	649	-----DVKYG-----NKGALVEYKVGTSMT-----	669
Qy	1423	NEGDR\$PQ\$SLV\$SPQADIPKVTQ\$EVH\$AQVEY\$KQ\$EVL\$ATV\$SGATP\$IVDEK\$PAPS\$VT	1482
Db	670	-----LP-ATWAE-----KVKAVLPL\$GICVRKP-----	692
Qy	1483	RGVKIDK\$KAVAHV\$AKQVQ\$VOP\$KOR\$SLTINEG\$KAGOLC\$OLCMFRTCSG\$VQ\$LDVYNEA	1542
Db	693	-----Q\$SKP-----LDEBDD-----	703

Qy	1543	TIATRFSA-FTFVDNLKGRSAVFVFSKLGEYTYNGGSHVSSGWPRPALEDILTAKTPSV	1601
Dz	704	--LRLSNMFFKYSDLK-----LKKTIIPVVVTGT	731
Qy	1602	FDHCLVKQYKMGGVPPHADDEECYPSDNPILITVLNKGANFSTCKRGKGVWVINVASG	1661
Dz	732	IREROMKNYI-----	741
Qy	1662	DYFLMPCCGFORTHLHSVNSIDEGRISLIIFRATRFRVGGVMQLAGGSDEKSJPGVPNQ	1721
Dz	742	DYLSASLGSLGNLERI-----	758
Qy	1722	PQSOGATRTITPKSGKALSEGSRGVKGRSTYSIW-CEDQYVRKCEWLADPNVMALEP	1780
Dz	759	-----VRSDWNGTE-ESMQTFLGYDCE-----KCKML-----LLP	787
Qy	1781	DYTPTFEVVKTGTSADVVEYLKYLAIGI---ERTYRALLMARNIAVTTAEGVLKVPN	1836
Dz	788	AEKKHAAVVLASDDTTRII-FLSYDESQPIIDKKNMKRFAVCSETKVYSVIRSLEVLN	846
Qy	1837	QVYESLGEHVYKSGTDLIHFHSQTODGLRVLDLPYVILAEKGIFFTQKDVAVALGDGNLF	1896
Dz	847	KEAIVDPGVHI-----TLVDG-----	863
Qy	1897	VCDAILVFHDAINLIGALKVARCGMVGESFKSFYKCYNAPPGGKTTLV-----DEF	1950
Dz	864	-----PGCKTAETIARVNWKDTL	882
Qy	1951	VKSPNSTATTANVGSSSEDINMAVKRDNLBLEGNSATNTSVRSVNFIV--RGMXY--RV	2006
Dz	883	VLTPGREAA-----AMIRRRACALHKSPVATNDNVRTDSFVMMNRKIFKFDVA	930
Qy	2007	LYDEVHMHHOGLLOLQVFGATCAGEGLIFGDIQIIPFINREKFVRMDCAVFPKK-----	2060
Dz	931	YVDEGLVAHTGLLNFAIKLSGCKKAFAVFQDAKIPIFN--VMNFD----YPKELRTLIV	984
Qy	2061	ESVYVTSKSYRCPLDVLCYLLSSMTVRGTEKCYPEKVSQDKDPVYRSLSK-----R	2111
Dz	985	DNVERRVTHCRPDVTSFUNT-----YKAAVATTSPPVHSVKATKVSAGILR	1034
Qy	2112	PIGTTDDVAEINADVLCMTQLEKSDMKRSLKGGKETPMVTMVEHAOGKTSDVVLFRTK	2171
Dz	1035	P-----ELTIKGGI-IITFTQSDQSLIKS-----CYNDVNVTVEHQGETFEETAUVRAT	1083
Qy	2172	KADDSLTCK-OPHILVGLSHRTSLRYAALSCLKDDKVGYTISDASPOSVS DALLHFTA	2229
Dz	1084	PTPIGLTARDSPHVIALTHRTKAMYTVT--VFDAVTSIADY--EKVOQSILTFEA	1137

RESULT 12

JQ1896  
 hypothetical 191.1K protein - raspberry bushy dwarf virus  
 C:Species: raspberry bushy dwarf virus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Mar-2000  
 C:Accession: JQ1896  
 C:Ziegler, A.; Natsuaki, T.; Mayo, M.A.; Jolly, C.A.; Murant, A.F.  
 J. Gen. Virol. 73, 3213-3218, 1992  
 A:Title: The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus.  
 A:Reference number: JQ1896; MUID:93107856  
 A:Accession: JQ1896  
 A:Molecule type: mRNA  
 A:Residues: 1-1694 <ZIE>  
 A:Cross-references: GB:S51557; NID:g262149; PIDN:AAB24599.1; PID:g262150  
 C:Genetics:  
 A:Map position: segment RNA-1  
 C:Superfamily: cucumber mosaic virus RNA 1 protein  
 C:Keywords: RNA replication

Query Match	2.5%	Score 291;	DB 2;	Length 1694;
Best Local Similarity	25.6%	Pred. NO. 1.6e-09;		
Matches 122;	Conservative	65;	Mismatches 186;	Indels 104; Gaps 15;

1

Db	644	VNIHFSTRGRMRPVEGIVISVGVNHRGLGPKSDGELYIVNSECVICNS-----ESLSTV	698
Qy	1912	GALKVARCGMVGESFKFEXKYNAPPGKGTITLTVDEFVKSPNSTATITIANVGSSDDIN	1971
Db	699	TRSLQAGTGTISQ-----VDGAVGCGKTTAIIKSIK--EPSTDMIVTANKKSAQDVR	747
Qy	1972	MAVKRKDPNLEGLNSATTNRSRVYFIVRGMVKRLVDEVMHMQGLQLGVFATGASEG	2031
Db	748	MALEPKSSDSREACTFVRTADSVLLNEC--PTVSRVLVDEVVLLHFGQLCAVMSKLLKAVRA	805
Qy	2032	LFPGDIHQIIPFINEKVFEMDCAVFVPKKSQVYTSKSRCPDLCVYLLSMTV-----RG	2087
Db	806	ICFGDSQIAFSSRSDASFDWRFSKIIIDPETSDAOT--TFRSPQDVPVLVRMTAKPALPKG	863
Qy	2088	TEKCYPEKVYSGDKPVPVRSLSKRPIGTTDDVAEINADVVICMTQLEKSDMKSLKGGK	2147
Db	864	THSKYTKWVSQSK---VKRSVTSRAIASVTLVDLDDLSRFVITMTQADKA-----SLNSRAK	916

Qy 2148 E-----TPVMTVHEAQKTFSDVWLFRTKKADBSLFTKQHILVLGSRHTRSLVY 2197  
- : |||:|| : | | | -|| : | | | : | | | : |  
Db 917 EMNLPKTFWNERIKTVHESGISEDHVTLVRLKSTKDLFKQFSYCLVALTRHKVKTRY 975

Search completed: January 31, 2001, 15:28:52  
Job time: 181 sec

RESULT 15

plvXpJ  
 N:1 protein - peanut stunt virus (strain J)  
 N:Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (E  
 C:Specics: peanut stunt virus  
 C:date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change '15-May-2000  
 C:Accession: JQ1451  
 R:Karasawa, A.; Kakutani, T.; Minobe, Y.; Ehara, Y.  
 J. Gen. Virol. 73, 701-707, 1992  
 A:Title: Nucleotide sequence analyses of peanut stunt cucumovirus RNAs 1 and 2.  
 A:Reference number: JQ1451; MUID:92185483  
 A:Accession: JQ1451  
 A:Molecule type: genomic RNA  
 A:Residues: 1-1005 <KAR>  
 A:Cross-references: DDBJ:D11126; DDBJ:D01123; NID:g222417; PID:g222418  
 C:Genetics:  
 A:Map position: scgment 1  
 C:Superfamily: cucumber mosaic virus RNA 1 protein  
 C:Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine  
 F:81-174/Domain: methyltransferase status predicted <MTP>  
 F:720-977/Domain: helicase status predicted <HHG>  
 F:722-729/Region: nucleotide-binding motif A (P-loop)

Query Match 2.4%; Score 275.5; DB 1; Length 1005;  
Best Local Similarity 24.8%; Pred. No. 6.1e-09;  
Matches 137; Conservative 86; Mismatches 229; Indels 101; Gaps 21;

[illegible]





DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-JUL-1989 (Rel. 11, Last annotation update)  
DE POTENTIAL 194 KDA PROTEIN (PUTATIVE REPLICASE) [CONTAINS: 134 KDA  
DE PROTEIN].  
OS Tobacco rattle virus (strain Sym).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88034943; PubMed=3668507;  
RA Hamilton W.D.O.; Boccaro M., Robinson D.J., Baulcombe D.C.;  
RT "The organisation and inter-viral homologues of genes at the 3' end of  
RT tobacco rattle virus RNA1.";  
RL EMBO J. 5:223-229(1986).  
RT J. Gen. Virol. 68:2563-2575(1987).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RA Boccaro M., Hamilton W.D.O., Baulcombe D.C.;  
RT "The organisation and inter-viral homologues of genes at the 3' end of  
RT tobacco rattle virus RNA1.";  
RL EMBO J. 5:223-229(1986).  
CC -!- SIMILARITY: THIS PROTEIN IS HOMOLOG TO PUTATIVE REPLICASE OF  
CC TOBACCO MOSAIC VIRUS (TMV) AND TRICORNAVIRUSES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X06172; CAA29537.1; -  
DR EMBL; D00155; BAA00110.1; -  
DR PIR; S01865; S01865.  
DR INTERPRO; IPR000606; -  
DR INTERPRO; IPR001788; -  
DR INTERPRO; IPR002588; -  
DR PFAM; PF00978; RNA\_dep\_RNAPol2.1.  
DR PFAM; PF01443; Viral\_helicase1.1.  
DR PFAM; PF01660; Vmethyltransf.1.  
FT CHAIN 1 1187 134 KDA PROTEIN.  
FT CHAIN 1 1707 194 KDA PROTEIN.  
SQ SEQUENCE 1707 AA; 194328 MW; F4DABDB6C7F51F04 CRC64;  
  
Query Match 2.9%; Score 338.5; DB 1; Length 1707;  
Best Local Similarity 18.0%; Pred. No. 1.9e-12;  
Matches 334; Conservative 209; Mismatches 514; Indels 801; Gaps 72;  
  
Qy 445 DAMIREKPS-----HRCDFVLPKPREKRLRELPELSITQFSDSVRSRSHPPFANAMRSCFN 498  
Db 50 DDMIREKTKDPVIVHE---VLSQEQNKLMIEYFEFNVFVKDDKNMVGFAAERKLOA 106  
Qy 499 GIFSRRCGNVC-FFDIGSFYHVKAGHNVCHVNPVLDVKDVKRRINEILFLSTAGGDS 557  
Db 107 LLLDRVPALQEVDDIGGQSFWRGKRIHSCCPNLDIRDQREISQIFLTAIGDQA 166  
Qy 558 YVSSDLLTE-----AASKSVSYCSRESQNCSDRADGF-----MV 592  
Db 167 RSGKQMSGENELMWDQPKKNIAPNAVRCNNTYGGCTCR---GFSQKKGGAQYAIALH 223  
Qy 593 DVIDSPQQAEMDKGALVFDAIMF-PVELLYGNGEVVLEELDTLVKREGDYLAENV 651  
Db 224 SLYDFPKLDMATWEKTKVVAHMLAPESMLVDEG-----PLPSVDGYTKKN- 274  
Qy 652 GCGGBMYBHSFSNVSGFF---TFYSVVRTSGNVFKLEYEGVRCGYHLLTMCRAQSPGTE 708  
Db 275 ---GKIY-----FGPEKDPFSYIHWE-----EYKKYLL-----CKP 304  
Qy 709 VTYRSLVPSFVGKSLVFLP-VVAGSSVSFKTIVLSDSFVDRIYSVALNTIGTFENRTFEY 767  
Db 305 VSYQ-----GNVFPEPQVRGDTMLFS-----IYRIA----- 332  
Qy 768 AVGAVRSQKTHVTGSRVHVSKVDISPDDMGLVVAVMAQAIDKRAKSIRSYNEIKASEG 827

Db 333 --GVPR-----RSLSSQEQYR----- 346  
Qy 828 SLAGVFKLFQTVGDCFSNAVSIVAKAMVDNPNVLETLMSPRAFIKRVPGSVVVITCT 887  
Db 347 -----RIVISRWNWVVPI-- 361  
Qy 888 SGASDRLELARGAFDISKETFGKRKLKNSRLRVFSRAIVDSIKVMKAMKTDGCKPLPITD 947  
Db 362 -----FDLVEST-----RELVKDLDLFEKQFM-----D 384  
Qy 948 SVYAFIMG-----NVSNVHCTRAGLLGGSKATVVSVSKGLVARGAATKAFSGITSFSS 1001  
Db 385 KCIDYIARLSDDOQLTISNV-----KSYLSSNNWVLFINGAAVK----- 422  
Qy 1002 TGSLFYDRCLTDERLDALVRT---ENAINSPVILETSRVAVSKVAGTKPEWSEVSLN 1058  
Db 423 -----NKQSVSDRIQLLAQTLLVKEQVAPV-----MRELREALTETKPI---TSLT 468  
Qy 1059 DFTTFVLRNKKVLIGIFVASLGAAPITAWKYRRGTAAANARYAGSSYETLSLSSQAAGGLR 1118  
Db 469 D-----VGLGISRKL-----WK-----QFA-----NKIAYGGFV 492  
Qy 1119 GLTSTVSGSLVVRGFSASVTVRATVAKRQVPLALLSFSYSIAISGCSGLIWAHAL 1178  
Db 493 GMV-----GTLL---GFYPKKVLT-----WAKDT 513  
Qy 1179 PHLMFFFGGLTLLGARASANTWKFQGSNNMCAPVEVVMRGKSVSSLLPITL-----G 1233  
Db 514 PNG-----PELCYENSHKTKVILSVVYVYIGG 541  
Qy 1234 VSLIIRGLLNDTIPOLAYVPVE-GRNVYD---ETLRYRDFDYDEGAGSGTQHEAVPGD 1290  
Db 542 ITLMRRDIRDGLVKKLCDFIKGAHVLDVENPCRY-----E 580  
Qy 1291 DNDGSTSSVSYDVVTNVRDVGISTNGEETHSPRSVQVTVVEEVAPSAVAERQ 1350  
Db 581 INDFSSLYSA-----SESGET-----VLPDLSVKKAK 608  
Qy 1351 GDPGSGTADAMAFVESVKKGVDDVPHQOSSGETAREVEVDGKGLLPESVWGEPATQERG 1410  
Db 609 SD-----KLLQKKEIADEFLSAKFSNYSGSSVRTS-----PPSVVGSRSRGLGL 653  
Qy 1411 RAADNTAQTAVNEGDRPEVQSSVSSPOADIPKVTQSEVHQAQEV-KQEVPLATVSGAT 1469  
Db 654 LLEDSNV-----LTQARVGVSRKVDDEIMEQFLSG-- 684  
Qy 1470 PIVDEKPAPSVTTRGVKLIIDKGAHAHVAEKKQOVQOPKQSLTINECKAGKQKCMFT 1529  
Db 685 -----LIDTEAIDEVVSASFACER-----GETSGTKVLCKPLT 719  
Qy 1530 CSCGVOLDVYNEATIAFRSNAFTFVDNLKGRSAVFFSKLGEQYTYNGGSHVSSGMPRAL 1589  
Db 720 -----PPGF 723  
Qy 1590 EDILTAIKYPSVFDHCLYOKYKMGVGPVHADDEECYPSDNPILTIVNLVGNKANSFTCKRK 1649  
Db 724 ENVLPAVK-----PLVSKGK-----TVKRV----- 743  
Qy 1650 GSKVMVINVASDGYFLMPCGF---QRTHLRSVNSIDEGRISLTFRATRRVFGVGRMLQLA 1706  
Db 744 -----DYQVMGGERLPRPVVSGNSVDARREFLYLLDAERVAQNDEIMSL- 790  
Qy 1707 GCVSDEKSPVNPQQPSOGATRTITPKSGGKALSBGSG---REVKGRTSYSTWCBDYV 1763  
Db 791 -----YRDYSRGVIRT-----GCQNYPHGLGVWDVEMKN-----WCIRPVV 826  
Qy 1764 RKEWLNRADNPVMALEPDYTPMTFEVVKTGTSEDAAVVEYLKYLAIQI-PRTYRALLMARN 1822  
Db 827 TEHAYV-----FQPD-----KRDMDWSGYLEAVAWER----- 853  
Qy 1823 IAVTTAEGVLKVPNOVYESLPGFHVYKSTDLIFHSTQDLRVRLDPYVLIAEKGIFTKG 1882  
Db 854 -----GMLVNDP----- 860

[illegible]



```
Db 579 TLAF--DFEFQDYCEKLG-----VDIVTVELD-AISTGRAG 614
QY 1318 ----EVTGEETHSPRSV--QYTYVEEVAAPSAAEROGDPSGSGTADAMAFVESYKGV 1372
Db 615 ITLDHTDKEEQPRTLACSSYLEEE-----PS-----643
QY 1373 DDVFHQSSGETAREVEVDGKLLPES-VVGEAPTOERGAADGNTAQTVAVNEGDEPVO 1431
Db 644 DDL-----VCLSDKAIVNRSTILGELN-----NVVIFEGTL-PKN 678
QY 1432 SSVLSPOADIPKVT--QSEVHAQKEVQEVPLATVSGATPLVD-----EKPAAPTTRGV 1485
Db 679 SVFVSAP--DDPSVIELSELHAR-----PVSDFLSMQKPNVIVTGEV 720
QY 1486 KIIDKGAVAHVAEKQVOVEQPKORSLTINEGKAGKQLCMFTCSGQVLDVYNEATIA 1545
Db 721 QI-----COMQNYLD-YLSASIV 737
QY 1546 TRFSNAFTVDNLKGRSAVFVSKLGEYTYNGSHVSSOWPRALEDILTAIKPSVFDHC 1605
Db 738 ACISNLKYLQDQWLNPEKFKQIGV-----WDNLNNKW-----IVVP-----775
QY 1606 LVQKYKMGGVFPFHADDECYPSPILTVNLVGRANFSTCKRKGKVMVINVASGDYFL 1665
Db 776 --QKKYANGLAADVGNQ-----KTVILNY-----799
QY 1666 MPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDKSPGVNPOQPSQ 1725
Db 800 -----HLDQ-----KKPTATITLVDGV-----850
QY 1726 GATRTITPKSGGKALSEGSGREVKCRSTYSIWCEODYVRKCEWLNRADNPVMALEPDYTPM 1785
Db 806 -----ILEKSYVRL-----814
QY 1786 TPEVVKTGTSDDAVVEYLKYLAIIGIERTYRALLMARNIAVTTAEGVLKVPQNVYESLPGF 1845
Db 815 ----VVSTDTYLFVVSMGLYLR-----833
QY 1846 HVYKSGTDLIFHSTQDGLRVDRDLYVLAEGAFTKGKDVAVALGDNLFVCDLILVPH 1905
Db 834 -----HLDQ-----KKPTATITLVDGV-----850
QY 1906 DAINLIGALKVARCCMVGESFSPKCYKNAPPGGKTTLLVDFPVKSPNSTATTANVG 1965
Db 851 -----PGCGKQTEILSRF--DANSDLIL---VQ 873
QY 1966 SSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVR-----GMVKRVLVDVHMHOGLLQL 2021
Db 874 GREACEMIRRRANDNVPG--SATKENVRTFDSFVNRKPGKKTLLWVDEGLMHPGLINF 931
QY 2022 GVFATGASEGLFFGDINOIPFINREKVRMDCAVFVPKKESVYVYTSKSYRCPLDVCYLLS 2081
Db 932 CINISCVSSVYIFGRKQIPIFINRVWNESIPDNLAKEYDEIVSRDRTTKRCPLDVTHFLN 991
QY 2082 S-----MTVRTEKCYEKVGVSGDKPKVRSLSRPTGTDDVAEINAD-VYLCMTQLE 2134
Db 992 SYVERVMSYSNVQRSLCKMTSGRAK-----INDYRSILAEGKLLFTFOED 1038
QY 2135 KSDMKRSLKGGKETPVMTVTHAEQCKTSDVVLFTKTKADDSLFTK-OPHILVGLSRHTR 2193
Db 1039 KEYL--LKAGFKD--VNTVTHAEQGETVRDVMNLI RTVATPLTIVSAGSPHTVVALSRHTN 1093
QY 2194 SLVYAALSSKLDKGYTGISDASPOSVSDDLHTFAPA 2231
Db 1094 RPVYVTV--VPDVVMTTVQ--KTCCVSNFLDMDTAVA 1126

RESULT 5
V125_AMYLE
ID V125_AMYLE STANDARD; PRT: 1126 AA.
AC P03589;
DT 21-JUL-1986 (rel. 01, Created)
```

```
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 01-NAR-1992 (rel. 21, Last annotation update)
DE 125 KDA PROTEIN.
OS Alfalfa mosaic virus (strain 425 / isolate Leiden).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Alfamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83143345; PubMed=6298738;
RA Cornelissen B.J.C., Brederode F.T., Moormann R.J.M., Bol J.F.;
RT "Complete nucleotide sequence of alfalfa mosaic virus RNA 1.";
RL Nucleic Acids Res. 11:1253-1265(1983).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00163; AAA46289.1;
DR PIR; A04197; WMEF12.
DR INTERPRO; IPR000606;
DR INTERPRO; IPR002588;
DR PFAM; PF01443; Viral_helicase1;
DR PFAM; PF01660; Vmethyltransf;
SQ SEQUENCE 1126 AA; 125827 MW; BF5A8019B47D4CBF CRC64;

Query Match 2.9%; Score 329.5; DB 1; Length 1126;
Best Local Similarity 16.9%; Pred. No. 3.4e-12;
Matches 318; Conservative 191; Mismatches 488; Indels 887; Gaps 63;

QY 435 TKLCDFMESO-----RDAMIREKPSHRCDF-----LKPREREKRLPELSIQFSDSV 483
Db 40 TAIGKVFSEAGRAYAQDAL-----PSDKGEVLKISFSLDATQONILRANPGRTRTFNSNS 95
QY 484 RSSHPFANAMRSCFNGIFSRRCGNV--CFDFDIGGSFTYHVKAGHVCHVCNPVLDVKDKV 541
Db 96 SSSHCFAAAHRLLETFDVFYRCFGNTVDSIIDLGGNFVSHMKVKNHNVHCCPILDARDGA 155
QY 542 RINEILFLSTAGDSYVSS--DLLTEAASKSVSYCSRESQNCDSRADAGFWV-DVYDTS 598
Db 156 RUTERILSLK-----SYVRKHPEIVGEA-----DYCMDTFQKCSRADYAFATHSTSDLD 205
QY 599 PQQVAEAMDKGALVFDIALMPVELLYGN-GEVYLEELDTLVKREGDVLAY-----NV 651
Db 206 VGLACSLDQKGMKFICTMVMVDAADMLIHNEGEIPNFVNRWEIDRKKDLIHDFIDEPNL 265
QY 652 GCGEMYEHFSFNSVSGFTFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGT---E 708
Db 266 G-----YSHRFSLLKHLTY-----NAVDLGHAAVR-----IERKQDFGGVWVID 305
QY 709 VTYRSL-----VPSFVGKSLVPIPVVAGSSV-----SFKTIVLDSDFVDRISY 752
Db 306 LTY-SILGFYKPMHPSNGRSCAWNRVYKGMVHTVNEGYHHSHSYQTAVRKKVLVDKVKLT 364
QY 753 ALNTGTGTENRTFEVAVGAVRSQKT-----HVTGSRVHVSKVDISPDMMGLVAVM 805
Db 365 RYTEVAFROFRPNADAHSAIQSIATMLSSSTNHTIIGVTLTISGKPLSPDDYIPVATTI 424
QY 806 AQAIAK-----DRAKSIRSYNFIKASEGSLAGVFKLFFQTVGCFNSNAVSVAK 853
Db 425 YRVKLYNNAIPEMLSLDKGERLSTDAVLKSGEGPM-----WYSGFTLSALD---K 473
QY 854 AMVHNFNVLETLMSPRAFIKKVPGSVVVVTICTSGASDRLELRGAFDTSKETFGKLN 913
Db 474 VNVPGDF-VAKALLSLPK-----RDLSK 495
QY 914 SRLRVFSRAIVEDSKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGSKA 973
Db 496 ----LFSRS-----ATSHSERTPVRDE-----SPIRC-----518
```









```
QY 1576 NGSVSGWSPRALDILTATKYPSVDFHCLVQYKMGVGFPHADDECPSPNPILTV 1635
   | : : : | : | : | : |
Db 720 -----AVSNLVKILKDTAAID--LETRQKEG----- 743
QY 1636 NLVGKANFSTCKRGGKGVWVNVASGDYFLMPC-----GFQTHLHSVN-----SIDEGR 1685
   | : : : | : | : | : |
Db 744 -----VLDVASKRWLVKPSAKNHAWGVVETHARKYHVHALLLEHDE-- 782
QY 1686 ISLFRATRRVFGVGRMLQLAGGVSDKSPGVNPQPOQOGATRTITPKSGGKALSEGSG 1745
   | : : : | : | : | : |
Db 783 -----FGI----- 785
QY 1746 REVKRSYISWCODYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSSEDAVVEYLKY 1805
   | : : | : | : | : |
Db 786 -----ITCD-----NWR--VAVSESVV---Y 803
QY 1806 LAIGIERTYRALLMARNIAVTAGCVLKVPNQVYESLPGFHVYKSGTDLIFHSQTODGLRV 1865
   | : | : | : : : | : |
Db 804 SDMAKLRTRLRKLKDGEPHVSSAKVVL----- 830
QY 1866 RDLPVVLIAGEGIFTKGRKDVDAVVALGDNLFVCDLILVFHDAINLIGALKVARCGMWGES 1925
   | : | : | : | : |
Db 831 -----VDGV----- 834
QY 1926 FKSFEYKCYNAPPGGKTTTLVDEFVKSPNSTATITANVGSSSEDINMA-----VKKRD 1978
   | : | : | : | : | : |
Db 835 -----PGCGK-----TKELSRVNFEEIDLILVPGRQAAMIRRR 868
QY 1979 PNLGCLNSATTVNSRVV-NFIV-----RGMVKRVLVDVEMHMQGLQLQGVFATGASEG 2031
   | : | : | : | : | : |
Db 869 ANASGIIVAKDNVTRVDSFLMNYCKGARCFKRLFIDEGIMLHTGCVNFIEMSLCDIA 928
QY 2032 LFFGDIQIPIFIREKVFEMDCAVFVKESVVTYSKRYCPDVCVLLSS-----MTV 2085
   | : | : | : | : | : |
Db 929 YVYGTOQIPINRVTCGPPYPAHFAKLEVDETRRTTLRCPADVTHFLNORYEGHVACT 988
QY 2086 RCTEKCEPEKYVSG--KDKPVYRSLSKRPICGTTDDVAEINADVYLCHMTQLEKSDMKRSLK 2143
   | : : : | : | : | : |
Db 989 SSEKKSQEMVSGAASINPVSFKLKGK-----ILTFQ-----SD-KEALL 1029
QY 2144 CKGKETPVMTVHEAQKTFSDVVLFRKKADDSLFTK-QPHILNGLSRHTRSLVYAA--- 2199
   | : | : | : | : | : |
Db 1030 SRG-YADVHTVHEVQGETYADVSLVRLTPTVPVSIARDSPHLVLSLRHTKSLKYTYVM 1088
QY 2200 --LSKSLDD--KVGTYISD 2214
   | : | : | : | : |
Db 1089 DPLVSIIRDLERSVSYLLD 1107

RESULT 9
RPO_TOMK1
ID RPO_TOMK1 STANDARD; PRT; 1616 AA.
AC Q901T8; Q901T7;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)]
OS Tobacco mosaic virus (strain Kazakh K1) (TOMV) (TMV strain K1).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE-20196905; PubMed-10732356;
RA Belenovich E.V., Novikov V.K., Zavrlev S.K.;
RT "Biological properties and genome structure of the Kazakh isolate K1
RL Mol. Biol. (Mosk) 34:172-176(2000).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
```

```
CC -----BETWEEN CODONS FOR GLN-1116 AND GLN-1118.-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).-----
CC EMBL; AJ243571; CAB62911.1; -
CC EMBL; AJ243571; CAB62912.1; -
CC INTERPRO; IPR000606; -
CC INTERPRO; IPR001788; -
CC INTERPRO; IPR002588; -
CC PFAM; PF00978; RNA_dep_RNapol12; 1.
CC PFAM; PF01443; Viral_helicase1; 1.
CC PFAM; PF01660; Vmethyltransf1; 1.
CC Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1616
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
FT SEQUENCE 1616 AA; 183606 MW; 4BP9913CE791DE33 CRC64;
SQ

Query Match 2.6%; Score 302; DB 1; Length 1616;
Best Local Similarity 16.7%; Pred. No. 3.le-10;
Matches 315; Conservative 204; Mismatches 516; Indels 848; Gaps 65;

QY 403 ALNGEEVYSSVSSITNRLVLRDQSALLSHDITKLCDFMSQORDAMIRKPSHRCDVFLKP 462
   | : : : | : | : | : |
Db 2 AYTQTATSSALLETVYRGNTLVNDLAKRRLYDTAV-DEFNARD---RRPKVNESKVYSE 56

QY 463 REREKRLPELSTQFSDSVSRSSHPFANAMRSC-FNGIFSRRCGNVCFDIGSGFTYHV 521
   | : : | : | : | : | : |
Db 57 EOTLIATKAYPEFOITFYNTQNAVHSLAGGRLSELEYLMQIIPYGSUTYDIGNFASHL 116

QY 522 KAGVNCVHCNPVLVDVKVRRINE---ILFLS--TAGGDSYVS-----SDLTE 566
   | : | : | : | : | : | : |
Db 117 FKGRAYVHCMPNLDVRDIMEHGEKQDSIELYLSRLERGNKHVPNFOKEAFDRYAEMPE 176

QY 567 AASHSVSYCSRESQNC-DSRADAGFMVDYDIPSOQVAEAMDKKALVFDIALMPFVELL 625
   | : | : | : | : | : | : |
Db 177 VVCHDTQTQCRHSQECYTGRTVYATIALHSYIDIPADEFCALLRNKRVHVCYAAAFSENL 236

QY 626 YNGEVLLEELDTLVKREGDYLVNVCQCGEM-YEHSPSNVSGGFTFSYVTRTSGNVFKL 684
   | : | : | : | : | : | : |
Db 237 LEDSHVNLDIEINACFQDGRDLTFSASESTLNTSHSYNLKVKCYTYFFASNREVYMK 296

QY 685 EYEGYRCGYHHLTMCRAQKSPCTEVTYRSLVPSFVGKSLVPIPVVAGSSVSEKTVILVSD 744
   | : | : | : | : | : |
Db 297 EF----- 298

QY 745 FVDRIYSY--ALNTIGTFENFTFEYAVGAVRSQKTHVITGSRVVHSHKVDISPDDMWGLV 802
   | : : | : | : | : | : |
Db 299 LVTRVNTWFCFESRIDTF-----LLYKGVAKHGVD----- 328

QY 803 AVMAQAIKDRAKRSIRSNFIKASEGLAGVFKLFPOTVGCDFCSNNAVSYAKAMVHDNFNV 862
   | : | : | : | : | : | : |
Db 329 -----NEQFYKAMEDAWH-----YKKTIAMCNSERILLEDSVVNYWEPKP 368

QY 863 LETLMSMPRAFIRKVPVGSVVVVTICTSGASDRLELRGAFDIDSKTFGRKLNKRSIRLVFSRA 922
   | : : | : | : | : | : |
Db 369 MRDMVIVP-----LFDISLET-----SKTRKEVL--- 393

QY 923 IVEDSI-KVMKAMKTEDEGKPLPITEDSVYATFMGNVSNVHCTRAGLGGSKATVVSSVSK 981
   | : : | : | : | : | : |
Db 394 VSKDFVTVLNHRTYQAKAL--TVSNVSEVESIRSRV-----IINGVTARSEWDYDK 445

QY 982 GLVARGATKAFSGITSPFSTGSLFYDRGLTEDERLDALVTENAINSPVGILETSRVAV 1041
   | : : | : | : | : | : |
Db 446 SLQSLSMT-----FF-----LHTKLAVLKDD-LLI 470
```

QY 1042 SKVVAGTK-----EFWSEVSLNDFTFVLRNKLVLIGIFVASLGAAPLAIANKYRRCIAANARR 1097  
DB 471 SFAFGPKTVSHVWDEISL----- 490  
QY 1098 YAGSSYETLSSLSOAGGLRGLTSTVSGGLVVRGFSAAVTVTRATVAKRQVPLALL 1157  
DB 491 ----- 490  
QY 1158 SFSTSYAISGCSMLGIWAHALPRHLMFFGLTGLGARASANTWKFSGSNMCAPVEV 1217  
DB 491 ----- 501  
QY 1218 WRGKSVSSLLPITLGVSLIRGLLNDRIPLQAVYVPPVEGRNVYDYLRYRDFDYDEGA 1277  
DB 502 ----RLINKLKII-----ENALEIRVPDL-YV----- 525  
QY 1278 GPSGTQHEAVPGDDNDGSTSSVSSDVVTNVRDVGISTNGEVTGEBEETHSPRSVOYTYVE 1337  
DB 526 ----TFHDLR-----VSEYKMSVDMVPVLDIRKKMEETEE-----MYNALS 561  
QY 1338 EVAPSAVAERQDPSGSGTADAMAFVESVKKGVDDVPHQOSSGETAREVEVDGKGLLP 1397  
DB 562 E-----LSVLKTSDFDVFDSQ-----CQSLEVD--PMTA 591  
QY 1398 ESVVGEAPTOERG-----RAADGNTAQTAVNEGDRPVOSSLVSPQADIPKVTOSEVH 1451  
DB 592 AKVIVAVMSNESGLTLTEQTEANVA-LALQDSEKASDGAIVVTSRDVEEPSIRGS--- 647  
QY 1452 AOKEVQEVPLATVSGATPIVDEKPAVPTTRGVRIIDKGAVAHVAEKKQVQEPQOR 1511  
DB 648 ---MARGELQAGLSG-----DVPESVY-----RSEIELESLEQF 679  
QY 1512 SLTINEGKAGOLCMFRCTSCGVQLDVNEATATFRSNAFTFVNLKGRSAVFFSKLGE 1571  
DB 680 HMATASSLIHKOMCSI-----VGTPLKVQOMKN---FIDSLVASLSA----- 719  
QY 1572 GYTYNGSHVSSGWPRALEDILTATKYSVPDHLVQYKMGVYFFHADDECYPSDNP 1631  
DB 720 -----AVSNLKILKDTAAID--LETRQKFG----- 743  
QY 1632 ILTVNLVAGKANFSTCKRGGKVMVINVASGDYFLMPC-----GFORTHLHSVN---SI 1681  
DB 744 -----VLDVASKRWLVKPSAKNHAWGVVEHARKYHVVALLEH 780  
QY 1682 DEGRISLTFRATRRVFGVGRMLQAGVSDKSPGVNPOQOSOGATRTITPKSGGKALS 1741  
DB 781 DE-----FGI----- 785  
QY 1742 EGSREVKGRTSYTWCEQDYVRKCEWLRAADNPVMALEPDYTPMTFVVVKTGTSADAVVE 1801  
DB 786 -----TCD-----NWRR-----VAVSSVYV- 802  
QY 1802 YLKLAIGIERTYRALLMARNIAVTTAEGVLKVPNOVYESLPGFHVYKSGTDLIFHSTQD 1861  
DB 803 ---ISDMAKRLTLRLKLDGEPHVSAAKVL----- 830  
QY 1862 GLVRDLPLYVLTAEKGIPTGKQDVAVALGDNLFVCDLILVFDHAINLIGALKVARCGM 1921  
DB 831 -----VDGV----- 834  
QY 1922 VGESKFSPEYKYNAPPGGKTTTLVDFEVKSPNSTAITANVGSSSEDINMA-----V 1974  
DB 835 -----PCGCK-----TKEILSRVNEEDLILVPGRQAAM 864  
QY 1975 KKRDPNLEGLNSATTNVRV-NEIV-----RGMVKVLDVDEHMHQGLQLQGVFATG 2027  
DB 865 IRRANASGIIVATKDNVTRVDSFLMNTYKGAQCQFKLFDIEGLMLHTGCVNFWLSL 924  
QY 2028 ASEGLFFGDIQIPINREKVFMDCAVEVPKESVWYTSKSYRCPDLVCYLLSS----- 2082  
DB 925 CDIAVYVGTQOIPIYINRVTPFPYPAHFAKLEVDEVEVTRRTILRCPADVTHFLNORYEGH 984  
QY 2083 -MTVRGTEKCYPEKVSVG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMK 2139

DB 985 VMCSTSEKSVSOEVMVGAAASINPSKPLGK-----ILFTTQ---SD-K 1025  
QY 2140 RSLKKGKGTPTVTHQAQKTFSDVLFRTKKADSLFTK-OPHILVGLSRHRSRLVYA 2198  
DB 1026 EALLSRG-YADVTHVEVOGETYADVSLVRLPTPTVSIITARDSPHVLVLSRHTKSLKY 1084  
QY 2199 A-----LSSKLLDD--KVGTYIISD 2214  
DB 1085 TVVMDPLVSIIRDLEVRVSSYLLD 1107  
RESULT 10  
RRPO\_CGMVS  
ID RRPO\_CGMVS STANDARD; PRT: 1648 AA.  
AC P19523; P89877; P90356;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:  
METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].  
OS Cucurbit green mottle mosaic virus (watermelon strain SH) (CGMV), and  
OS Cucurbit green mottle mosaic virus (watermelon strain W) (CGMW),  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
RN [1]  
SEQUENCE FROM N.A.  
RP STRAIN-SH;  
RX MEDLINE=91311400; PubMed=1856687;  
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,  
RA Sato T., Motoyoshi F., Nishiguchi M.;  
RT "The complete nucleotide sequence of cucumber green mottle mosaic  
virus (SH strain) genomic RNA.";  
RL J. Gen. Virol. 72:1487-1495(1991).  
RN [2]  
RP REVISIONS;  
RC STRAIN-SH;  
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,  
RA Sato T., Motoyoshi F., Nishiguchi M.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE OF 1496-1648 FROM N.A.  
RC STRAIN-W;  
RX MEDLINE=89073773; PubMed=3201760;  
RA Saito T., Imai Y., Meshi T., Okada Y.;  
RT "Interviral homologues of the 30K proteins of tobamoviruses.";  
RL Virology 167:653-656(1988).  
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL  
RNA REPLICATION.  
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA  
CAPPING AND AN RNA HELICASE.  
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN  
CODONS FOR LYS-1144 AND GLN-1146.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D12505; BAA18895.1; -  
DR EMBL; B12505; BAA18896.1; -  
DR EMBL; J04322; AAA46382.1; -  
DR PIR; J01157; WMTWS2.  
DR INTERPRO; IPR000606; -  
DR INTERPRO; IPR001788; -  
DR INTERPRO; IPR002588; -  
DR PFAM; PF00978; RNA\_dep\_RNAPol2; 1.  
DR PFAM; PF01443; Viral\_helicase1; 1.  
DR PFAM; PF01660; Vmethyltransf; 1.  
KW Transferase; 1 1648 RNA-DIRECTED RNA polymerase; Helicase; ATP-binding.  
FT CHAIN



FT CHAIN 1 1144 METHYLTRANSFERASE/RNA HELICASE.  
FT NP\_BIND 863 870 ATP (POTENTIAL).  
SQ SEQUENCE 1648 AA; 186549 MW; 1DIAFPEER7B65595 CRC64;

Query Match 2.68; Score 298; DB 1; Length 1648;  
Best Local Similarity 17.4%; Pred. No. 5,7e-10;  
Matches 334; Conservative 192; Mismatches 509; Indels 884; Gaps 72;

QY 405 NOREYVSS---VVSSTNRLRLDQDQALLSHLDKICDMFSQORDAMIREKPSHRCDVEL 460  
DB 10 NORDAASGRNLVQSQASKRYDEAVRSLDQD-----RPKMNFGRV 54  
QY 461 KPREREKRLPELSIQFSDSVSRSHPFANAMSCFNGIFSRRC--GNVCFEFDIGSFT 518  
DB 55 STEHRLVTDAYPEFSISFTATKNSVSHSLAGLRLLELYMMQVPYGSPC-YDIGNYT 113  
QY 519 YHVKAGHNVCHVNPVLVDKVKRRI--NEILF-----LSTAGDSYVSSD 562  
DB 114 QHLEKGRSYVHCCNPNCLDLKDVARVMYNDMITQHVQRHKGGGCRPLPTFQIDAFRRYD 173  
QY 563 LLTEAASKSVSYCSRESQNCDSRADG-----PMVDVYDISPOVAEAMDKKALVFDI 616  
DB 174 -----SSPCAVTCSDFQECSDYDFGSGRDHNAVLSHIYDIPYSSIGPALHRKNVRVCYA 228  
QY 617 ALMPPEVLLXNGEYVLEELDTLVKREGDYIAYNVGQCGEM-YEHSFSNVSGFFTFYSVR 675  
DB 229 AFHFSEALLLGSVGNLNSIGAOPRVGDVHFLFSEESTLHYTHSLENIKLIWMRYFP 288  
QY 676 TSSGNVFKLEYGRGCHYHLLTCRA-----OKSPG-----TEVTVRSVPVSGSL- 723  
DB 289 ADDRFEVYKEFMVKRVDTFFRLVRADTHMLHKSVGHSKSKSYFALNTPIPFQDKATF 348  
QY 724 -----VFIPV-----VAGSVSEKFTIVLDSDFVDRIYSVALNTIGTFENR--TFE 766  
DB 349 SVWPEAKRVKLVLPKFELSFELSGNVKISRLMVDADVFHTI-----INHSTYDNKALVWK 404  
QY 767 YAVGAVRSQKTHVITGSRVHRSKVDISPDMMGLVAVMAQAIAKDRAKSIR--SYNFIKA 824  
DB 405 NVQSFSVESIRSVLVNGSVKSEWNPVVDQTDISFIFL-LVKRVKQVIELMSDKVIE 463  
QY 825 SEGSL---AGVFKLFFQTGDCFSNNAVSVYAKAMVHNFNVLETLMSPRAFIRKVPGSV 881  
DB 464 ARGLLRRFADSLKSASVGLGDC-----VY----- 487  
QY 882 VVTICTSGADRLRLGAFDTSKETFGKRLKNSRLRVFSRAIVDSIKVMKAMKTEDGKP 941  
DB 488 -----DALVQTGWFTS-----SDEKVL-----LP 508  
QY 942 LPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVYSSVSKGLVARGAATKAFSGITSFSS 1001  
DB 509 EPPWTFSDYL-----EGMYEADAKIERESVSELLASGDDLFKKIDIRNNYS 555  
QY 1002 TGSFYDRGTEDERLDALVRTENAINSPGILETSRVAVSKVVGAKFVSEVSLNDET 1061  
DB 556 -----GVEFD-----VEKFEQEFCKELNV----- 573  
QY 1062 TFVLNKNVLGIFVASILGAAPIAWKYRRGIAANARRVAGSSYETLSLSQAAAGLRLGT 1121  
DB 574 -----NPMWLG-----HVIEAIFSKAG----- 591  
QY 1122 SSTVSGSLVVRGFSVAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRH 1181  
DB 592 -----VIVT----- 595  
QY 1182 LMFFFGGLTGLGARASANTWKFGFSNMCAVPEVWRGKSVSLLLPITLGVSLIIRGL 1241  
DB 596 -----GLGTL-----SPEN-----GASVA----- 609  
QY 1242 LNDTIPQLAYVPPVEGRNVYDETLRYRDFYDEGAGPSGTQHEAVPGDDNDGSTSVSS 1301  
DB 610 LSST----- 620

QY 1302 YDVVTNVRDVGISTNGEVTGEEETHSPRSQYQTYTVEEVAPSAVAERQSGSPSGTADA 1361  
DB 621 MDVTEDMEDI-----VLMADKSHS-----YMSPEMARWA----- 649  
QY 1362 MAFVESYKKGVDYVPHQOSSGETAREVEVDCKGLLPESVVGCEAPT-----QERGRAADGN 1416  
DB 650 -----DVKYG-----NNKGALVEYKVGTSMTPLPATWAEKKA--- 682  
QY 1417 TAQTAVNEGDRPQYSSLVSSPQADIPKVTQSEVHAQVEKQVPLATVSGATPIVDKPK 1476  
DB 683 -----LPLSGICVRKPQ-----FSKPIDEE-----DDLIR 706  
QY 1477 APSVTRGVKIIDKGAHAVHAEKKQVQEQPKQRSILTINEGKAGKQLCMFRTCSGVQL 1536  
DB 707 LSNMFEKVSCLKLKTTPVVTGTIREROMK----- 739  
QY 1537 DVEYNEATIRFSNAFTVDMLKGRSAVFFSKLGEYTYNGSHVSSGWPRALEDILTAI 1596  
DB 740 -----NYIDYL-----SASLGSTLG-----NLERIVRSWNGTEESMOTFG 775  
QY 1597 KYPSVEDHCLVQYKMGGGVPPHADDERCYPSDNPILLTVNLVGRANFSTCKRGGKVMVI 1656  
DB 776 LY-----DCEKC----- 782  
QY 1657 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVRMLQLAGGVSDKSPG 1716  
DB 783 -----KWLLELPA--EKKHAWA-----VLASDDTTRIF--LSYDES-G 816  
QY 1717 VPNOQPOSOGATRITPKSGGKALSEBSGREVKGRSTYSIWCEODYVRKCCWLADNPVM 1776  
DB 817 SP-----IIDKNNKRFALVCSETKV-----YSV-----IRSEVL----- 846  
QY 1777 ALEPDYTPMTEVVKVGTGTSADVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPN 1836  
DB 847 -----NKEAIVD----- 853  
QY 1837 QVYESLPGFHYKSGTDLIFHSTODGLRVLDPYVILAEKGIPTKGDDVAVVALGDNLF 1896  
DB 854 -----PGVHI-----TLVDGV----- 864  
QY 1897 VCDLILPHDINLIGALKVARCGMVGESFKSPYKCNAPPGGGKTTTLV-----DEF 1950  
DB 865 -----PGCGKTAETIIARVNWKTDL 883  
QY 1951 VKSPNSTATITANVGSSBDINNAVKKRDPNLEGLNSATTVNSRVVNFV--RGMVK--RV 2006  
DB 884 VLTGREA-----AMIRRRACALHKSPVATNDNVRTFDSFVMNRKIKFKDAV 931  
QY 2007 LVDEVHMMHQLLQLGVFATGASEGLFFGDIQIPFINREKVFMRMDCAVFVPKK----- 2060  
DB 932 YVDEGLMVHTGLLNFALKISCKKRAVFGDAKQIPFINR--VMNFD-----YPKELRTLIV 985  
QY 2061 ESVVYTSKSYPCPLDVCYLLSSMTVVRGTEKCYPEKVSQKDKPVVRSLSK-----R 2111  
DB 986 DNVERRYVTHRCPRDVTSLNTI-----YKAAVATTSPVHVSKAIKVSAGILR 1035  
QY 2112 PIGTTDDVAENADVYLCMTOLEKSDMKRSLKKGKETPVMTVHEAGCKTSDVVLFRPK 2171  
DB 1036 P-----ELTKIKGKI--TFTTOSDKSLIKS-----GYNDVNTVHEIGCTEETAVVRAT 1084  
QY 2172 KADDSLETK-OPHILVLGSLRHSRLVLAALSCKLDDKVGTYISDASPOSQSDALIHTEA 2229  
DB 1085 PTPIGLIARDSPHVLVALTRHTKAMVYTV---VFDATVTSIADV--EKVDQSILTMFA 1138  
RESULT 11  
RRPO\_PMMVJ ID RRPO\_PMMVJ STANDARD; PRT; 1612 AA.  
AC P89657: p90347;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) (CONTAINS:



METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].  
 Pepper mild mottle virus (strain Japan) (PMV-J).  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
 [1]  
 SEQUENCE FROM N.A.  
 Kirita M., Akutsu K., Watanabe Y., Tsuda S.;  
 "Nucleotide sequence of the Japanese isolate of pepper mild mottle  
 tobamovirus.";  
 Ann. Phytopathol. Soc. Jpn. 0:0-0(1997).  
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL  
 RNA REPLICATION.  
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA  
 CAPPING AND AN RNA HELICASE.  
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS  
 BETWEEN CODONS FOR GLN-1117 AND GLN-1119.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AB000709; BAA19167.1; -;  
 DR EMBL; AB000709; BAA19166.1; -;  
 DR INTERPRO: IPR000606; -;  
 DR INTERPRO: IPR001788; -;  
 DR INTERPRO: IPR002588; -;  
 DR PFAM; PF00978; RNA\_dep\_RNAPol2; 1.  
 DR PFAM; PF01443; Viral\_Helicase1; 1.  
 DR PFAM; PF01660; Vmethyltransf; 1.  
 KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
 FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.  
 FT CHAIN 1 1117 METHYLTRANSFERASE/RNA HELICASE.  
 FT NP\_BIND 835 842 ATP (POTENTIAL).  
 SQ SEQUENCE 1612 AA, 183303 MW, E7E66CE8D81FE4C CRC64:

Query Match 2.6%; Score 297; DB 1; Length 1612;  
Best Local Similarity 22.8%; Pred. No. 6.3e-10;  
Matches 195; Conservative 114; Mismatches 301; Indels 24

[illegible]

Db	732	-----DTAAIDLETKEFGV-----YDVCLIKLWLKPLSKGHWGVVM- 769
Qy	1891	LGDNLFVCDDILVFDHAINLIGALKVARCG-----MVGESEKSEFYKCY----- 1934
Db	770	--DSDYKCFVALLIYDGENIV-----CGETWRRVAVSSSESLVSDMGKIRATRSVLKD 820
Qy	1935	NAPPGGGKTTLLVDFVKSPNSTATITANVGSSSEDINMAVKK-----RDPNLEGLNSA 1987
Db	821	GEPISSAKVTLVDG-VPGCGKTKELLSRVNFDDELVLVPGKQAAEMIRRRANSSGLIVA 879
Qy	1988	TTVNSRVV-NFIV--RG--MYKRVLVDEVHMHQGLLQLVFATGASEGLFFGDIQIIP 2041
Db	880	TKERVIVTDFSEFLMYNGRGPQYKRLFLDEGLMLHPGCVNFLVGLMSLCSEAEFVYGDQIQIP 939
Qy	2042	FINEKVFMRMCVAFVPPKKESVWVTSKSYRCPLOVCYLLSS-----MTVRGTCKCYPEK 2095
Db	940	YINRVATFPYPKHLISQLEVDVATVETRTTLRCPADITFFLNQKYEGQVMCTSSVTRKSVSHE 999
Qy	2096	VVSGKDKPVRVRSLSKRPTIGTTDDVAENADVYLWMTQLEKSDMKRSLKGGKGETPVMTVH 2155
Db	1000	VIQG--AAVMNPVSKPLKGKGV-----ITFTQSDKS-----LLLSRGVE-DVHTVH 1041
Qy	2156	EAGKTSDDVVLFRTKRADDLSFTKQ-PHILVGLSRHTRSLVY-----AALSSKLDK-- 2207
Db	1042	EVQGETEDVSLVRLTPTPVGIISKQSPHLLVLSLRHSIKYITVTVLDVAVSVLRDLEC 1101
Qy	2208	VGTVISDASPQSUS 2221
Db	1102	VSYLLDMYKVDVS 1115
RESULT 12		
RPO_PPMVS		
ID	RPO_PPMVS	STANDARD; PRT; 1612 RA.
AC	P29098; Q84924;	
DT	01-DEC-1992 (Rel. 24, Created)	
DT	01-OCT-2000 (Rel. 40, Last sequence update)	
DT	01-OCT-2000 (Rel. 40, Last annotation update)	
DE	RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:	
DE	METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].	
OS	Pepper mild mottle virus (strain Spain) (PMMV-S).	
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.	
RN	[1]	
RX	SEQUENCE FROM N.A.	
RA	MEDLINE=92113528; PubMed=1765765;	
RA	Alonso E., Garcia-Luque I., de la Cruz A., Wicke B.,	
RA	Avila-Rincon M.J., Serra M.T., Castresana C., Diaz-Ruiz J.R.:	
RT	"Nucleotide sequence of the genomic RNA of pepper mild mottle virus,	
RT	a resistance-breaking tobamovirus in pepper."	
RL	J. Gen. Virol. 72:2875-2884(1991).	
CC	-1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL	
CC	RNA REPLICATION.	
CC	-1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA	
CC	CAPPING AND AN RNA HELICASE.	
CC	-1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS	
CC	BETWEEN CODONS FOR GLN-1117 AND GLN-1119.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .	
CC	-----	
CC	EMBL; M81413; AAB02334.1; -	
CC	EMBL; M81413; AAB02335.1; -	
DR	PIR; J01312; WMTMPV.	
DR	INTERPRO; IPR000606; -	
DR	INTERPRO; IPR001786; -	
DR	INTERPRO; IPR002588; -	
DR	PFAM; PF00978; RNA_dep_RNapol12; 1.	
DR	PFAM; PF01443; Viral_helicase1; 1.	



QY 685 EYEGYRCGYHLLTMCRAQKSPGTEVYRSLVPSFVGKSLVFTPVVAGSSVSEKTIIVLDSO 744  
 Db 297 EF-----YHVALLEYDEQGIIVTCDDWRR-----VAVSESVW-- 802  
 QY 745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSOKTHVITGSRVHVHVKDISPDDMWGLVY 802  
 Db 299 LVTRVNTWPKFSRDTF-----LLYKGVAH----- 324  
 QY 803 AYNAQAIDRAKISIRSYNFIKASEGLAGVFLFOTVGDGCFNSNAVSVYAKAMVHDFNV 862  
 Db 325 -----KSDVSEQYITAMEDAWH-----YKTLAMCNSERILLDSSSVNVWFPK 368  
 QY 863 LETLMSMPRAFIRKVPGVVVVITICTSGASDRLELRGAFDISKETFGKLNKSLRVFSRA 922  
 Db 369 MRDWIIP-----LFDISLET-----SKSRKEVL--- 393  
 QY 923 IVEDSI-KVMKAMKEDGKPLPITEDSVYAFIMGNVSNVHCSTRAGLGGSKATVYSSVSK 981  
 Db 394 VSKDFVFTVLNIRTYQAKAL--TVANVLSFVESIRSV-----IINGVTARSEMDVDK 445  
 QY 982 GLVARGAATKAFSGITSPFSTGSLFYDRGLTDERLDALVRTEINAINSPVGLIETSRVAV 1041  
 Db 446 SLQSLSMT-----FF-----LHTKLAVLKDD-LLI 470  
 QY 1042 SKVWAGTK-----EPWSEVSL-----NDFTTEV--LRNKVLIGIEVASLGAAPIAWKYRRGTA 1092  
 Db 471 SKFSLGSKTVCHVDEISLAFGNAPPSVKERLLNKKLI----- 509  
 QY 1093 ANARYAGSSYETLSSLSQAAAGLGLTSTSVSGSLVVRGFSFSAVTVTRATVAKROV 1152  
 Db 510 -----RAAGDALEI-----KV 520  
 QY 1153 PLLALSFSTSYAISGSMGLGWAHALPHLMFFFGGLTLLGARASANTWKFGFSNNWCA 1212  
 Db 521 PDLYITF----- 527  
 QY 1213 VPEVWVRGKSVSLLLPITLGVSLIIRGLLNDTIPQAYVPPVEGRNVYDETLRYVRFD 1272  
 Db 528 -----HDLRVAEYK----- 536  
 QY 1273 YDEGAGSGTOHEAVPGDDNDGTSVSSYDVVTVNVRDVGISTNGEVTGEETHSPRSVQ 1332  
 Db 537 -----SSVDMPALDIRKM-----EET-----EYM 556  
 QY 1333 YTYVEEVAPSAABRQDPSGSGTADMAVESVKKGVDDVFHQSSGETAREVEVDG 1392  
 Db 557 YNALSE-----LSYLRESKDFDVFDSQM-----CKSLEVD- 587  
 QY 1393 KGLLPESVYGEAPTORG-----RAADGNTAQATVNEGDEPREPVQSSLSVSPQADIPKYT 1446  
 Db 588 -PMTAAKIVAVMNSESGLTLTFERTEANVAQ-ALQDOEKASEGALVTVSREVEEPSMK 645  
 QY 1447 QSEVHAQKEVQOEVLATVSGATPIVDEKPAFVSVTTRGVKIIDKGKAVAHAEKQOVQE 1506  
 Db 646 GS-----MARCELQAGFAGDHP-----ESSYSRNEEIE 674  
 QY 1507 QPKQSLTINEKAGKQLCMFTCSGQVOLDVYNEATATATRSNAFTFVDNLKGRSAVFF 1566  
 Db 675 SLEQFHMATADSLIRKOMSSI-----VYTGPIKVVQMKN-----FIDSLVASLSA- 719  
 QY 1567 SKLGEYTYNGSGHVSNGWPRALEDILTAIKYPSVFDHCLVOKYKMGVGPVPHADDECY 1626  
 Db 720 -----AVSNMLIKLDTAID--LETRQKFG----- 743  
 QY 1627 PSDNPILTIVNLGKANFTKCRKGKVMVINVASGDYFLMPCGQFQTHLHVSNSIDEGRI 1686  
 Db 744 -----VLDVASRKWLKP----- 756  
 QY 1687 SLTFRATRRVGVGRMLQAGVSEKSPGVNQPQSQOGATRTITPKSGGKALSEGSR 1746  
 Db 757 -----TAKSHANGVVVETHAR----- 771  
 QY 1747 EVKGRSTYSI-----WCODYVRKCEWLADNPVMALEPDYTPMTFEVVKVTGTSEDAAVVEY 1802

Db 772 K-----YHVALLEYDEQGIIVTCDDWRR-----VAVSESVW-- 802  
 QY 1803 LKYLAIIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHYKSGTDLIFHSTQDG 1862  
 Db 803 --YSDMAKLRTLRLLRDGEPHYSNKAVVL----- 830  
 QY 1863 LVRVDLPYVLLIAEKGIFTTKGDKVDVAVVALGDNLFVCDLILVFDHAINLIGALKVARGCMV 1922  
 Db 831 -----VDGV----- 834  
 QY 1923 GSEFSKFEYKYNAPPGGGKTTTLVDVEFKVSPNSTATITANVGSSEDINMAVKK----- 1976  
 Db 835 -----PGCGK-----THEILSRVNFDEDLILVCPKQAAEMI 865  
 QY 1977 -RDPNLEGNLSAT-----TVNSRVNF--IVRGMVKRLVDEVHMHQGLQLGLGVFATGA 2028  
 Db 866 RRANSSGIIIVATKDNVTVTVDSEPMNFGKTRCOFKRLFIDEGLMLHTGCVNFLVAMSLC 925  
 QY 2029 SEGLFPGDINQPIFNREKVFMDCAVFPKESVYVTSKSYRCPLDVCYLLSS----- 2082  
 Db 926 DVAYVYGDTOQIPIYINRVSGFPYPAHFSKLEVDEVETRTTLRCPADVTHYLNRRYEGFV 985  
 QY 2083 MTVRGTEKCYPEKVSQKDPVVRSLSKRPITGTTDDVAEINADVYLCMTOLEKSDMKRSL 2142  
 Db 986 VSTSSVKKSQVSGMWSG--AAVINPISKPLHGK-----ILTFQ--SD-KEAL 1028  
 QY 2143 KGKGKTPYTVHEAOGKTFSDVVLFRKKADDSLEF-KOPHILVGLSRHRSRLVAAA-- 2199  
 Db 1029 LSRG-YSEVHTVHEVQGEYISDVSLVRLTPTPISIIAGDSPHVLVALSRHTCSLKYYTVV 1087  
 QY 2200 ---LSKLLDD--KVGYYISD 2214  
 Db 1088 MDPLVSIIRLEKLSYLLD 1107  
 RESULT 14  
 VIA\_BMW STANDARD; PRT; 966 AA.  
 AC Q00020;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE LA PROTEIN [INCLUDES: HELICASE; METHYLTRANSFERASE].  
 OS Broad bean mottle virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;  
 OC Bromovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BANDEN;  
 RX MEDLINE-92074218; PubMed-1962437;  
 RA Dzanott A.M., Bujarski J.J.;  
 RT "The nucleotide sequence and genome organization of the RNA-1 segment  
 in two bromoviruses: broad bean mottle virus and cowpea chlorotic  
 mottle virus".  
 RL Virology 185:553-562(1991).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE REPLICATION OF THE VIRUS.  
 CC CONTAINS AN HELICASE DOMAIN AND A METHYLTRANSFERASE DOMAIN. THE  
 CC METHYLTRANSFERASE DOMAIN IS PROBABLY INVOLVED IN VIRAL RNA  
 CC CAPPING.  
 CC -!- SIMILARITY: TO IA PROTEIN FROM CCMV, CMV, PSV AND TAV.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: M65138; AAA42740.1; --  
 DR PIR: A41699; P1BVB8.  
 DR INTERPRO: IPR000606; --

```
DR INTERPRO: IPR002588; -.
DR PFAM: PF01443; RNA_dep_RNapol2; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
KW Helicase; ATP-binding; Transferase; Methyltransferase.
FT NP_BIND 690 697 ATP (POTENTIAL)
SQ SEQUENCE 966 AA; 109621 MW; DF592681D7231C8D CRC64;

Query Match      2.5%; Score 286.5; DB 1; Length 966;
Best Local Similarity 30.9%; Pred. No. 1.2e-09;
Matches 105; Conservative 46; Mismatches 124; Indels 65; Gaps 13;

Qy 1920 GWSGESKSFY-----KCNAP-----PGGKTTTLVDFV 1951
| :| | :| | :| | | | | | | | | | | | | | | | |
Db 646 GGLGEKFEFNLVVDKTCSCANAKAIADACKKVSAPTCSVVVMVDGVAGCGKTTAIKETF- 704

Qy 1952 KSPNSTATITANVGSSDINNAAVKRDPNLEGLNSATTVNSRVV-----NFIVRGMYK--R 2005
| :| | | | | | | | | | | | | | | | | | | | | |
Db 705 -RFEKDIIVTANRKSADVRKAI-----FGDASDSEVALKVVRTADSAIMHGLPECHR 756

Qy 2006 VLVDVHMHGQLQLGVFATGASGLFFGDIINOIPFINREKVRMD-CAVFEVPKESVW 2064
| :| | | | | | | | | | | | | | | | | | | | | |
Db 757 LLVDEAGLLHYGQLLAVADLCKSEVLAFGDEQLSFKSRDATFRMKVCNIEYDKRDIV- 815

Qy 2065 YTSKSYRCPLDVCYLLSMVTVRGTEK-----YPEKVVSGDKRPVNSLSKRPIGTDD 2118
| :| | | | | | | | | | | | | | | | | | | | | |
Db 816 --SKTFRCPDQV--VSAVKIL-KRKCANRSKNGWVSSK---VEKSLSKSRIVSINQ 866

Qy 2119 VAEINADVYLCTQLERSDKRSLLKGG-----KETPWTYVHEAQGTFSDVVLFRPKKA 2173
| :| | | | | | | | | | | | | | | | | | | | | |
Db 867 VSMKEHKFYLMTADKAAKLSRAKDVGLDKTWVESNNETVHEAQGKAVDHHVLVRLKST 926

Qy 2174 DDSLFTKPHILVGLSRHSRLVVAALSSKLDKDKVGYIS 2213
| :| | | | | | | | | | | | | | | | | | | | | |
Db 927 KCDLFKSEYCLVALTRHKRTFEYLYNGCDLGLDSIFVT 966

RESULT 15
RRPO_TMVB STANDARD; PRT; 1616 AA.
AC O9YQ08; O9WKD4;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobacco mosaic virus (strain B935A) (TMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Xue C., Zhou X., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence and genome organization of tobacco
RT mosaic virus isolated from Vicia faba.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011933; CAA09875.1; -.
DR EMBL: AJ011933; CAA09874.1; -.
DR INTERPRO: IPR000606; -.
DR INTERPRO: IPR001788; -.

```

```
DR INTERPRO: IPR002588; -.
DR PFAM: PF00978; RNA_dep_RNapol2; 1.
DR PFAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183296 MW; 99A63B441290B6CF CRC64;

Query Match      2.5%; Score 285; DB 1; Length 1616;
Best Local Similarity 16.9%; Pred. No. 3.5e-09;
Matches 318; Conservative 196; Mismatches 525; Indels 840; Gaps 62;

Qy 403 ALNDEEVSVSVSSITNRLVLRDQSALLSHDLTKLCDFMSQORDAMIREKPSHRCDFVLPK 462
| :| | | | | | | | | | | | | | | | | | | | | |
Db 2 AYTOTATTALLDTVTVRGNNSLVNDLAKRLLDYDVA-EEFNARD-----RPKVNFSKVIS 56

Qy 463 REREKRLRELFPELSIQFSDSVSRSSHPFANMRSC-FNGIFSRRCGNVCFFDIDIGGSFTYHV 521
| :| | | | | | | | | | | | | | | | | | | | | |
Db 57 EQTLIATRAYPEFQITFYNTQNAVHSLAGLSLELYLMMQIPYGSILTYDIGGNFASHL 116

Qy 522 KAGHVNHCVNPVLDVKDKRRINE-----ILFLS--TAGGDSY-----VSSDLLTEAASK 570
| :| | | | | | | | | | | | | | | | | | | | | |
Db 117 FKGRAYVHCCMPNLDVRDMRHEGQKDSIELYLSRLERGKTVPNFQKEAFDPAEYAEIPED 176

Qy 571 SVSVCRESQNC-----SRADAGFMDVVDYDISPOQVAEAMDKKALGVFDIALMFPVE 623
| :| | | | | | | | | | | | | | | | | | | | | |
Db 177 AV-CHNTFTCEHPQMOSGRVYAIALHSIYDIPADEFGAALLRKNVHTCYAAAFHSEN 234

Qy 624 LLYNGEYVLELDTLVKREGDYLAYNVGQCCEM-YEHSFSNVSGFTFSYVVRTSSGNVF 682
| :| | | | | | | | | | | | | | | | | | | | | |
Db 235 LLEDSEYVNLDEINACFSRDGDKLTFSPASESTLYNCHSYSNILKYCKTFPPASNREYV 294

Qy 683 KLEYEGYRCGYHLLTMCRQAQSPGTEVTYRSVLPSFVGKSLVFIPIVWAGSSVFKTI 742
| :| | | | | | | | | | | | | | | | | | | | | |
Db 295 MKEF----- 298

Qy 743 SDFVDRIYSY--ALNTICTGFENRFEYAVGAVRSOKTHVITGSRVSVHSHKVDISPDMMGL 800
| :| | | | | | | | | | | | | | | | | | | | | |
Db 299 --LVTRVNTWCKFSRIDTF-----LLYKGVAH----- 324

Qy 801 VVAVMAQAIDKRAKSIRSYNFIKASEGLAGVFKLFFOTVDCFSNAYSVYAKAMVHDF 860
| :| | | | | | | | | | | | | | | | | | | | | |
Db 325 -----KSVDSQQFYTAMEDAH-----YKTLAMCNSERILLEDSSSVNWF 366

Qy 861 NVLETLSMPRAFIRKVPGVVVVITCTSGASDRLELRGAFDISKETFGKRLKN---SRLR 917
| :| | | | | | | | | | | | | | | | | | | | | |
Db 367 PKMRDMVIVP-----LFDISLETSKTRKEVLVSKDF 398

Qy 918 VFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLGGSKATVVS 977
| :| | | | | | | | | | | | | | | | | | | | | |
Db 399 VFT-----VLNHIIRTYQAKAL--TYANVLSFVESIRSrv-----IINGVTARSEW 441

Qy 978 SVSKGLVARGAATRAFAFGITSFSTGSLFYDRGLTDEDERLDALVRTENAINSPVGISETS 1037
| :| | | | | | | | | | | | | | | | | | | | | |
Db 442 DYDKSLIQ-----SLSMTFY-----LHTKLAVLKDD 467

Qy 1038 RVAVSKVVAGTK-----EFWSEVSLNDFTTFVLNRNKLIGIFVASLGAAPIAWKYRRGIAA 1093
| :| | | | | | | | | | | | | | | | | | | | | |
Db 468 -LLISKFSGLSKTCVQHVDSEISL----- 490

Qy 1094 NARRYAGSSYETLSLSLSSQAAGGLRGLTSSVTSGGSLVVRGFSsavIVTRATVAKRQVP 1153
| :| | | | | | | | | | | | | | | | | | | | | |
Db 491 ----- 490

Qy 1154 LALLSFSTSYAISGCSMLGIWAHALPHLMFFFGTLGARASANTWKFQGSNNWCAV 1213
| :| | | | | | | | | | | | | | | | | | | | | |
Db 491 -----AFGNAPFSV 499

Qy 1214 PEVWNRGKSVSLLLPITLGVSLIIRGLLNDTIPOLAYVPPVEGRNVYDETLRYRYDFDY 1273
| :| | | | | | | | | | | | | | | | | | | | | |

```

Job time: 404 sec

```
Db 500 KERLLNRKLR-----VAGDALEIR-----VPDL-VV----- 525
QY 1274 DEGACPGSTQHEAVRGDDNDGSTSSVSYDVYNTNVRDVGISTNGEVTGEEETHSPRSVOY 1333
Db 526 -----TFHDL-----VTEYKASVMDPALDIRKKMEET-----EVMY 557
QY 1334 TYVEEVAPSAAVAERQDPSGSGTADAMAFVESYKGVDDVFHQSSGETAREVEVDGK 1393
Db 558 NALSE-----LSVLRSDKFDVDVFSQ-----CKSLEVD-- 587
QY 1394 GLLPESVVGAEAPTQBRG-----RAADGNTAQAVNEGDRPVSLSVSSPOADIPKVTQ 1447
Db 588 PMTAAKVIIVADMNSELGLTLPETERTEANVA-LALQDQEKASEGALVVTISREVEPSPKMG 646
QY 1448 SEVHAQKEVKQEVPLATVSGATPIVDEKPAVPTTRGVKIIDKRAVAHVAEKKOVQVBEQ 1507
Db 647 S-----MARGELQLAGLADHP-----ESSYSKNEETES 675
QY 1508 PKORSITINEGKAGKOLCMFRCTSCGVQLDVYNEATIRFSNAFTFVDNLKGRSAVFFS 1567
Db 676 LEQFHMATADSLIRKOMSSI-----VYTGPIKVOOMKN---FIDSLVASLSA--- 719
QY 1568 KLGEYTYNGGSHVSSGMPRALEDILTALIKPSVDFDHCIVOKYKMGGVVPFHADDEECYP 1627
Db 720 -----AVSNLVKILKDTAAID--LETRQKFG----- 743
QY 1628 SDNPILTIVNLVGKANFSTCKRGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRIS 1687
Db 744 -----VLDVASRKWLTKP----- 756
QY 1688 LTRATRVRVGVGRMLQLAGVSDKSPGVNQPSQCATRTITPKSGGKALSEGSGRE 1747
Db 757 -----TAKSHANGVVEVTHARK 772
QY 1748 VKGRSTYSI-----WCEQDYVRKCEMLRADNPVMALEPDYTPMTFEVVKVTGTSEDAAVEYL 1803
Db 773 -----YHVALLEYDEQGVVTCDDRR-----VANSSSEVV-- 802
QY 1804 KYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHYKSGTDLIFHSTQDGL 1863
Db 803 -YSDMAKLRTLRLRLRNGEPHVSSAKVVL----- 830
QY 1864 RVYRDLPYVLLAEBKGIPTKGKDYDAVVALGDNLFCVDDILVFHDAINLIGALKVARCGMVG 1923
Db 831 -----VDGV----- 834
QY 1924 ESFKSFYKCYNAPPGGKRTTLVDFEVKSPNSTATITANVGSSSEDINNAVKK----- 1976
Db 835 -----PGCGK-----TKEILSRVNFDEDLILVPGKQAAEMIR 866
QY 1977 RDPNLEGLNSAT-----TVNSRVVNF--IVROMYKRVLVDEVHMMHQGLLQIGVPATCAS 2029
Db 867 RRANSGIIVATKDNVKTVDSPMNF GKSTRCQFKRLFIDEGLMLHTGCVNPLVAMSICE 926
QY 2030 EGLFTGDIQIPFINREKVFMDCAVFPKESVVYTSKSYRCPLDVCYLLSS-----M 2083
Db 927 IAYVYGDYQQIPYINRVSGFPYPAHFALVEVDETRRTTLRCPADVTHYLNRRYEGFVM 986
QY 2084 TVRGTEKVCYPEKVVSGDKPVVRSLSKRPIGTTDDVAEINADVILCMTOLEKSDMKRSLK 2143
Db 987 STSSVKKSVSQEMVGG--AAVINPISKPSHGK-----ILFTQ---SD-KEALL 1029
QY 2144 GKGETPVWTVHEAOGKTFSDVVLFRTKKADDSLEF-KOPHILVGLSRHRTSLVYAA--- 2199
Db 1030 SRG-YSDVHTVHEVOGETYSVDSLVRLTPTPVSIAGDSPHLVALSRHTCSLKYYTVVM 1088
QY 2200 --LSSKLDD--KVGTYSID 2214
Db 1089 DPLVSIIRDLEKLSYLSD 1107
```

Search completed: January 31, 2001, 15:34:29

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:32:37 ; Search time 102.04 Seconds  
(without alignments)  
2569.523 Million cell updates/sec

Title: US-09-301-906-15  
Perfect score: 11520  
Sequence: 1 MDVIRPLRVSPFHVNNTLE.....QSVSDALLHTFAPAGCGRGI 2237

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_15:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertibrate:\*
  - 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3644	31.6	701	12	071188
2	803.5	7.0	398	12	09Q604
3	794.5	6.9	2473	12	071209
4	741	6.4	3090	12	09Q703
5	736	6.4	3070	12	089906
6	734	6.4	3124	12	066237
7	721	6.3	2301	12	040959
8	721	6.3	3074	12	008534
9	706.5	6.1	3115	12	09WID7
10	689.5	6.0	3115	12	09IFX0
11	678.5	5.9	3115	12	010467
12	661.5	5.7	3107	12	087587
13	643	5.6	1873	12	083044
14	483.5	4.2	1334	12	09JGK3
15	466	4.0	1350	12	09WIE3
16	466	4.0	1858	12	09WJD7
17	453.5	3.9	1345	12	09QBU8
18	453.5	3.9	1853	12	09IWA9
19	450.5	3.9	1270	12	093163

20	450.5	3.9	1778	12	056790	056790	beet soil-b
21	448	3.9	1277	12	09JGJ7	Q9JGJ7	sorghum chl
22	447.5	3.9	1320	12	Q06359	Q06359	soil-borne
23	447.5	3.9	1828	12	Q89249	Q89249	soil-borne
24	437.5	3.8	1308	12	Q09090	Q09090	soil-borne
25	437.5	3.8	1816	12	Q9JAD4	Q9JAD4	soil-borne
26	436.5	3.8	1306	12	Q9WK10	Q9WK10	broad bean
27	436.5	3.8	1820	12	Q9YPH5	Q9YPH5	broad bean
28	427.5	3.7	1308	12	Q09006	Q09006	soil-borne
29	427.5	3.7	1816	12	Q9JAD2	Q9JAD2	soil-borne
30	423.5	3.7	1308	12	Q9QCE8	Q9QCE8	European wh
31	423.5	3.7	1816	12	Q9IWB1	Q9IWB1	European wh
32	422.5	3.7	1308	12	Q09008	Q09008	soil-borne
33	422.5	3.7	1816	12	Q9JAD3	Q9JAD3	soil-borne
34	413	3.6	1302	12	Q9QBU0	Q9QBU0	potato mop-
35	391	3.4	1707	12	Q06806	Q06806	grapevine t
36	389	3.4	1707	12	Q67704	Q67704	grapevine v
37	386.5	3.4	710	12	Q65853	Q65853	beet yellow
38	373	3.2	1310	12	Q9WK22	Q9WK22	beet virus
39	373	3.2	1818	12	Q09099	Q09099	beet virus
40	359	3.1	2038	12	Q9IH80	Q9IH80	cherry necr
41	344	3.0	1180	12	Q84919	Q84919	pepper ring
42	344	3.0	1698	12	Q84918	Q84918	pepper ring
43	338.5	2.9	1188	12	P90434	P90434	tobacco rat
44	337	2.9	1130	12	P89202	P89202	sunhemp mo
45	336.5	2.9	1707	12	056250	056250	tobacco rat

ALIGNMENTS

RESULT 1

071188 PRELIMINARY: PRT: 701 AA.  
ID 071188  
AC 071188: TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 77.0 KDA PROTEIN (FRAGMENT).  
OS grapevine leafroll-associated virus 3.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;  
OC Closterovirus  
OX NCBI\_TaxID=55951;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NY1;  
RX MEDLINE=98264508; PubMed=9603346;  
RA Ling K.-S., Zhu H.-Y., Drong R.F., Slightom J.L., McPerson J.R.,  
RA Consalves D.;  
RT "Nucleotide sequence of the 3'-terminal two-thirds of the grapevine  
RT leafroll-associated virus-3 genome reveals a typical monopartite  
RT closterovirus.";  
RL J. Gen. Virol. 79:1299-1307(1998).  
DR EMBL: AF037268; AAC40717.1; -.  
DR INTERPRO: IPR000606; -.  
DR PFAM: PF01443; Viral\_helicase1; 1.  
FT Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 701 AA; 77008 MW; 772A764D2CA2A2DC CRC64;

Query Match 31.6%; Score 3644; DB 12; Length 701;  
Best Local Similarity 100.0%; Pred. No. 3.8e-248;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1537 DYVNEATATATRESNAFTFVDNLKRSVAFPSKLGEGTYNGSHVSSGWPRALEDILTAI 1596  
|||||

DB 1 DYVNEATATATRESNAFTFVDNLKRSVAFPSKLGEGTYNGSHVSSGWPRALEDILTAI 60  
|||||

QY 1597 KYPVSFVDFHCLVOKYMGKGVPFHADDEECYPSDNPILITVNLVKANFSTKCRGGKVMYI 1656  
|||||

DB 61 KYPVSFVDFHCLVOKYMGKGVPFHADDEECYPSDNPILITVNLVKANFSTKCRGGKVMYI 120  
|||||

```
Qy 1657 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTERATRRVFGVRMLQLAGGVSDKSPG 1716
Db 121 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTERATRRVFGVRMLQLAGGVSDKSPG 180
Qy 1717 VPNOQPOSOQATRTITPKSGKALSEGSGREVKGSRSTYSINCEODYVRKCEWLADNPVM 1776
Db 181 VPNOQPOSOQATRTITPKSGKALSEGSGREVKGSRSTYSINCEODYVRKCEWLADNPVM 240
Qy 1777 ALEPDYTPMTTEVVKVGTGTSADVVEYLYKLAIGIERTYRALLMARNIAVTTAEGVLKVPN 1836
Db 241 ALEPDYTPMTTEVVKVGTGTSADVVEYLYKLAIGIERTYRALLMARNIAVTTAEGVLKVPN 300
Qy 1837 QVYESLPGFHYKSGTDLIFHSTQDGLRVRLDPLVLLAEKGIPTKGDVDVAVGALGNLF 1896
Db 301 QVYESLPGFHYKSGTDLIFHSTQDGLRVRLDPLVLLAEKGIPTKGDVDVAVGALGNLF 360
Qy 1897 VCDLILVFDHAINLIGALKVARCMGVGESFKSFYKCYNAPPGGKTTTLVDFEVKSPNS 1956
Db 361 VCDLILVFDHAINLIGALKVARCMGVGESFKSFYKCYNAPPGGKTTTLVDFEVKSPNS 420
Qy 1957 TATTANVGSSSEDINMAVKRDPNLEGLNSATTVNSRVVNFIVRGMVKRVLVDEVHMHQ 2016
Db 421 TATTANVGSSSEDINMAVKRDPNLEGLNSATTVNSRVVNFIVRGMVKRVLVDEVHMHQ 480
Qy 2017 GLIQLGVFATGASGLFPGDINQIPFINREKVFPRMDCAVFVPKESVYVTSKSYRCPLDV 2076
Db 481 GLIQLGVFATGASGLFPGDINQIPFINREKVFPRMDCAVFVPKESVYVTSKSYRCPLDV 540
Qy 2077 CYLLSSMTVRGTEKCYPEKVYSGDKPVPVRSLSKRPIGTTDDVAEINADVLCMTQLEKS 2136
Db 541 CYLLSSMTVRGTEKCYPEKVYSGDKPVPVRSLSKRPIGTTDDVAEINADVLCMTQLEKS 600
Qy 2137 DMKRSKLGKGTETPVMTVHEAOGKTFSDVVLFRTRKKADDSLFTKQPHILVGLSRHTRSLV 2196
Db 601 DMKRSKLGKGTETPVMTVHEAOGKTFSDVVLFRTRKKADDSLFTKQPHILVGLSRHTRSLV 660
Qy 2197 YAAISSKLDKVGTYIISDASQSVSDALLHTFAPAGCFRGI 2237
Db 661 YAAISSKLDKVGTYIISDASQSVSDALLHTFAPAGCFRGI 701

RESULT 2
Q90604 PRELIMINARY; PRT; 398 AA.
AC Q90604;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE HELICASE (FRAGMENT).
GN HEL.
OS grapevine leafroll-associated virus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=47985;
RN [1]
RP SEQUENCE FROM N.A.
RA Fazeli C.F., Rezaiian M.A.;
RT "Nucleotide sequence and organization of ten open reading frames of
RT the grapevine leafroll-associated virus 1 genome and identification of
RT three subgenomic RNAs.";
RL J. Gen. Virol. 0:0-0(2000).
DR EMBL; AF195822; AAF22737.1; -.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicase1; 1.
KW Helicase.
FT NON_TER
SQ SEQUENCE 398 AA; 44426 MW; A908109D49C22A23 CRC64;
```

Query Match 7.0%; Score 803.5; DB 12; Length 398;  
Best Local Similarity 44.2%; Pred. No. 3.8e-48;  
Matches 174; Conservative 59; Mismatches 156; Indels 5; Gaps 2;

```
Qy 1847 VYKSGTDLIFHSTQDGLRVRLDPLVLLAEKGIPTKGDVDVAVGALGNLFVCDLILVPHD 1906
Db 3 VYDANFKVYVANAPEDGTTIFDLEYVFLVSTGTETVPRRLQAVLSRQDAVLVCDLILVPHD 62
Qy 1907 AINLIGALKVARCMGVGESFKSFYKCYNAPPGGKTTTLVDFEVKSPNSTATITANVGS 1966
Db 63 AMNLRGCVRLAKRAMVGEYMKDVRISAVNSPPGGGNTTRLVDEYFGRKKRAKIAAANTGS 122
Qy 1967 SEDINMAVKRDPNLEGLNSATTVNSRVVNFIVRGMVKRVLVDEVHMHQGLLQGVFAT 2026
Db 123 VADINAAIRAREGKKEPDLVAKTANSVWVNSHPNSHVLGIDVYMLHKGFMQLTVVSM 182
Qy 2027 CASGLFPGDINQIPFINREKVFPRMDCAVFVPKESVYVTSKSYRCPLDVDCYLSSMTVR 2086
Db 183 GVKEVIAYGDKNQIPFINREKTFVTPNEAVEFAEEQIDYTDISYRCPADVCLVSSMTDA 242
Qy 2087 GTEKCYPEKVYSGDKPVPVRSLSKRPIGTTDDVAEINADVLCMTQLEKSPMKRSLKG-- 2144
Db 243 RGMKMPNGVPGNDVRPLRSFEKVIATPEDALLYEADVYLTMTQNEAKEMQKAVAKME 302
Qy 2145 --GKETP--VMTVHEAOGKTFSDVVLFRTRKKADDSLFTKQPHILVGLSRHTRSLVYAALS 2201
Db 303 VVAGKKRPDVTTHAEAGKTYENVVLRLKADDPIFSRRKPHIVVALSRHTRSMKYAVLS 362
Qy 2202 SKLDDKVGTYIISDASQSVSDALLHTFAPAGCFR 2235
Db 363 SKMTDTISKLDGTSAGKVSVDLLQLQQRNDRFR 396

RESULT 3
O71209 PRELIMINARY; PRT; 2473 AA.
AC O71209;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE METHYLTRANSFERASE/HELICASE POLYPROTEIN (FRAGMENT).
OS grapevine leafroll-associated virus 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=64003;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98264507; PubMed=9603345;
RA Zhu H.Y., Ling K.S., Gosczyński D.E., McFerson J.R., Gonsalves D.;
RT "Nucleotide sequence and genome organization of grapevine leafroll-
RT associated virus-2 are similar to beet yellows virus, the
RT closterovirus type member.";
RL J. Gen. Virol. 79:1289-1298(1998).
DR EMBL; AF039204; AAC40855.1; -.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Methyltransf.1.
KW Transferase; Methyltransferase; Helicase; Polyprotein.
FT NON_TER
SQ SEQUENCE 2473 AA; 277096 MW; F8DEBA6A58ED425 CRC64;
```

Query Match 6.9%; Score 794.5; DB 12; Length 2473;  
Best Local Similarity 21.6%; Pred. No. 6.5e-46;  
Matches 538; Conservative 342; Mismatches 913; Indels 699; Gaps 113;

```
Qy 152 QDEQKPKRQASHWAVKPTAVGVHVPKPKQEALEP-AQSVPPQQL-----EKAAL 201
Db 158 QSSVYKTKREDAHRTVEERAAGSVQEPQKRIDEKCGCRVPSGGFSLHVLNLEVRKV 217
Qy 202 TFGLIFFSKGGGDESADVILRKGLFNALNVPIDVKNFTFWAK-----IWDEASRRR 253
Db 218 AAGLLRFRVGGDM-----FHRSFSTQAG-HRLLVWRRSSRSCVLELYSPSKNFL 266
Qy 254 GYFVVK---DRAVFFPPIVRCRATIEDFIVNTAPGCDVALPRIELWSRERAFVCTYKGW 310
Db 254 GYFVVK---DRAVFFPPIVRCRATIEDFIVNTAPGCDVALPRIELWSRERAFVCTYKGW 310
```



Db 267 RYDVLPCSDYAAAMPFAAGGRFPL-----VLMTRIRY-----PNGF 303  
QY 311 CWFNNERLGEIYRRRCFSSSISG-----FLMHLGFRSLKURIRFAGTNTLHMP 359  
Db 304 CYLAHCYACAFLLRGDFPKREDIGAFPTAAKLNRNWSSELGERSL----- 349  
QY 360 SLNEERTFCWKG-----GDVYLPNVPKTAIVAGDRTRLGCEILASVANALNOEB 408  
Db 350 GUNLYGAYTSRGVHCYDAKFIKURL-----MSAVIAG-KDGEVVPSPDITPAMKQK- 403  
QY 409 YSVSVSSITNRLRLDQOALLSHLDTKLCDFMSQDAMIRESKPSHRCDFVFLKPREREL 468  
Db 404 ----TIEAVYDLRGCTDLSLLKSLTEKOLIDEKNDVQSLKDKRPIVKVPFYHSEATQNSL 459  
QY 469 RELFPELSLTOFSDSVSRSHIPPANMRSCENGIFSRRCGNVCFDIDGGSTYHVKA-GHVN 527  
Db 460 TRFYQFELFKSHSHSDHPAAASRLLENETLVLRLCGN-SVSDIGGCPFLHLHSKTQRR 518  
QY 528 CHVCNPVLVDKVKRR-INEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCDSRA 586  
Db 519 VHVCRPVLGDKAQRVRVRLDQYSNVRLGD----DDKILE-GPRNIDICHYPLGACDHES 573  
QY 587 DAGFMVVDYDISPOQVAEAMDKGALVFDIALMFPVELLYGNGEYVLEELDTLVKRE--G 644  
Db 574 SAMMVQVYDASLYEICGAMIKKSRITVLTWTPGCELDGRECVYMESLDCEIEVDVHA 633  
QY 645 DYLAYNVGCGMEYHSFNSGFFTSYVRTSSGNVFKLEYEGYRCGYGHHLTMCRAOKS 704  
Db 634 DVVMYKFG--SSCYSHKLSIIRKIDMTTPYL-TLGGFLFSVMYEVNRMGYFKITKSEVS 690  
QY 705 PCTEVT-----YRSL-----VPSVKGSLVFIPIVWAGSSVSFKTIVLDSFVDRIYSY 752  
Db 691 PSISCTKLURYRANSDDVVKLPFRDKRRMCLP-----GYDPIYLDOSKFSRVFDY 743  
QY 753 ALNTICTFEIRTFEXAVGAVRSQKTHVITGRSVHVSKVDISPDDMMGLVAVVMAOAIAKDR 812  
Db 744 VVNCNSAVNSKTFEWWWSIFKSKSRVLIISGLKIIHKNVNDLKYVESFAVMMLASCVRSP 803  
QY 813 -----AKSIRYNFIKASEGLAGVKLFOTQVGDG-FSNVAVSYVAKAMVHD---NFNV 862  
Db 804 LASEYAKNLSHFS-----GDCSFEATSFVLRKIRKNTMTNFN- 842  
QY 863 LETLASMPR--AFIRKVPGSVVVTICTSGASDRLELRGAFDISK-----ETFGRLKNS 914  
Db 843 -ERLQOLVKRVAFATLDVSLDLDSTLESITDFAECKVAIELELDELGLCLRAEAENKIRN- 900  
QY 915 RLRFVSRAIVEDSIKVMKAMKTE--DGKPLPTEDSVYAFIMGNVSNVHC-----TRAG 966  
Db 901 -----LAGDSIAAKLASEIVVDIDSKPSKQ-----VGNSSSENADKREVORPG 944  
QY 967 LLGSGKATVSVSVKGLVARGAATKAFSGITSFSTGSLFYDRGLTEDERLDALVRTENA 1026  
Db 945 LRGRSRNGVUGE-----FLHFVVDLSALRLFKYATDQORIKSYVR---F 984  
QY 1027 INSPVCILETSRVAVSKVAVGTFEWS-----EVSNDFTT 1062  
Db 985 LDSAVSFIDYNDLUSFLIRVLESGVSCMFAFLNRGDLSSRVSRAVCAKEVATSCANA 1044  
QY 1063 FYLRNKLVLIGIEVASLGAIPAIAMKYRR-GIAANARRYAG-----SSYE 1104  
Db 1045 SVSKAKVMI-----TFAAVAVCAMMENSFCGSGDREYKSIHYRTOVLFTDITFFEDSSYL 1099  
QY 1105 TLSSLSQAAGLRLTSS-----TVSGSLVVRGRFSSAVTVTRATVAKRQVPLALLSFS 1160  
Db 1100 PIEVLSSAICGAIVTLFSSGSSISLNAFLQITKGFSEVVY-----RNVVRVTHGLS 1152  
QY 1161 TSYAISGCSMLGIWAHALPHLM-----FFFCGLTLLGARASANTW-----KF 1203  
Db 1153 TT-ATDGV-IRGVFSQ-IVSHLLVGNVTGNVAYOSAFIAGVWPLLVKKCVSLIFILREDTY 1209  
QY 1204 GGFSSNMCAVPEVW-----RGKSVSSLLPLITLGVSLIIRGL----- 1241  
Db 1210 SCGFIKH--GISEFSLSSILKFLKLVDEL-----KSIQGVFDSNKKHVFKEATQE 1259

QY 1242 -LNDTIPO--LAYVPPVE--GRNVYDETLRYRDFDYDEGAGSPGTOHEAVPGDNDG- 1294  
Db 1260 AIRTTVMQVPVAVDALKSAAGRIYNNFTSR--RTEKDEGSSSDGACEEYFSCDEGEP 1317  
QY 1295 STSSVSSYD-----VVTNVRDVGIStNGEVTGEEETHSPRSVQYTYVVEEVAAPSAA 1345  
Db 1318 GLKGGSYGFSILAFFSRIMWGARRLIVKVKHECFGR-----LFEFL-----SLK 1362  
QY 1346 VAERQDPSGSGTADAMAEVSVYKKGVD-----VFHQSSGSETAREVEVDGKGLLPESV 1400  
Db 1363 LHEERTRVSGKNRTDV-----GVYDFLPTGIVETLSSIEECQIE-----ELGDDL 1409  
QY 1401 VGEAPTOERGRAADGNTAQTAVNEG-REPVOSSLVSSPOADIPKVTQS----- 1448  
Db 1410 KG-----DKDASLTDNMNYFSEDFLASIEEPFAGLRGSGKNAIILAILEYAH 1458  
QY 1449 ---EVHAQKEVKEVPLATVSGATPIVDEKPAVSPTTRGVKIIDKGKAVAHAEKQOVQ 1505  
Db 1459 NLFIRIVASKCRKPLFLAFELSSALIEK--FKEVPR-----KSOLVAIVREYTO-RF 1509  
QY 1506 EQPKORSLTINEKAGKQLC-----MFTCSGCVQOLDVNE-----ATIAFRFSNAF 1552  
Db 1510 LRSRMALGLNNEFVVKSPADLLPALMKRKVSGSELASVYRPLRGFSYMCVSAERREKFF 1569  
QY 1553 TFDVNLKGRSAVFFSK-----LGEYF----- 1574  
Db 1570 ALV-CLIGLSLPPFVRIVGAKACEELVSSARREYERIKIFLRQKYVVSLSNFFCHLFSSDV 1628  
QY 1575 -----YNGGSH-----VSSGMPRALEDILTAKYPSVFDHC--LVOK 1609  
Db 1629 DSSASAGLKGASRMTLPHLLVRLASALLSGW-----EGLKULLSHHNLFLCFALVDD 1684  
QY 1610 ---YKMGGVFPFHADDECYPSPDNPILTV-----NLVGKANFSTCKRK 1649  
Db 1685 VNVLIKVLGSLFFV-----QPIFSLFAAMLLQDPDFEYSEKLYTAPFEFFLKCS 1735  
QY 1650 GKGVMVINVASGDYFLMPCGFORTHLSVNSIDEGISLITFRATRVF-----GVGRML 1703  
Db 1736 RAPALL-----KGF-----FECVANSTVSKTVRRLRCFVKMLKLRGRGL 1776  
QY 1704 OLAG-CVSDEKS-PGVPNQPOSOGATRITPKSGKALS-----ESGRE 1747  
Db 1777 RADGRGLHROKAPVPIPSNRVVTGVERLSVKMQGVEALRTELURILEDLDSAVIEKLNRR 1836  
QY 1748 VKGRSTYSICWEOQYVRK-----CEMLRADNPVMALE-----PDYTPM 1785  
Db 1837 -RNROTN---DDEFTPAHEQMOEVTFCs--KANSAGLALERAVLVEDAIKSEKLSKT 1889  
QY 1786 TFEVVKTG--TSEDAVV-----EYLYLAIGIER-----TYR-----ALLMARNI 1823  
Db 1890 VNEMVRKSGTTEBVAVALSDDEAVEEISVADERDDSPKTVRISEYLNRLNLSSEFFPKPI 1949  
QY 1824 AVTTAGVLKVPNOVYESLPGFHVYKSGTDLIFHST---ODGLRV-----RDL 1868  
Db 1950 VDDNKDTGGLTNAVRE-----FYMOELALFEITHSKLCTYYDOLRIVNDRSVAPCEA 2005  
QY 1869 PYVLTAEKGIPTKGKDVA-VVALGDNLFVCCDIL-----VFHDAINLIGALKV 1916  
Db 2006 QLYVVRKNGSTIVQGEVRLHINDPHDHFDFGKISINKRRRCNGVLYHD--NL--AFLA 2061  
QY 1917 AKGMVGESEF-----KSFEYKCYNAPGGGKTTTLVDEEV-----KSPNSTATITANV 1964  
Db 2062 SNLFLAGYFPRSFRFTNSSVDILLYEAPGGGKTTTLIDSLFKVFKKEGVSMILTANK 2121  
QY 1965 GSSEDINMAVKRDPNLE-----GLNSATTNVNRVNVFVGRMYKRVL-VDEVHM 2014  
Db 2122 SSOVELKLVKEVSEVSNIECOKRKRSPKSIYTDAYLMHH--RGCDADVLFIDCFMW 2179  
QY 2015 HOGILQGLVFATGASGLFFGDINQIPFINR---EKVFRMDCAVFPKKEVSVYTSKSYR 2071  
Db 2180 HAGSVLACIEFTRCHKVMTIFGDSRQIHYTERNELDKCLYDLDLRFV-DLOCRVYGNISYR 2238

Qy 2072 CPLDVCYLLSSM-----TVRGTEKCYPEKVVVSGKDKPVRVSLSKRPPIGTTDDVAEINAD 2125  
Db 2239 CPWDVCAWLSVYGNLIATVAGESE-----GKS-----SMRINEINSVDDLDVPDVG 2285  
Qy 2126 VYLCMTQLEKSDMKRSLKGG-KETPVMTVHQAOGKTFSDVVLPRTKKADDSLTTPKOPHI 2184  
Db 2286 TFLCMLQSEKLEIKHFIRKGLTKLVNLTVEHAQETIYARVNLVRLKQEDPEPKSIRHI 2345  
Qy 2185 LVGSRHTRSLVYAAALSKLDDKVGTYISDAS 2216  
Db 2346 TVALSRRHTDSLTYNVLAARRCDATCDALQKAA 2377  
RESULT 4  
Q0703 PRELIMINARY; PRT; 3090 AA.  
AC Q0703;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
DE P348.  
OS Beet yellows virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;  
OC Closterovirus.  
OX NCBI\_TaxID-12161;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BV-4;  
RX MEDLINE-20079557; PubMed-10611288;  
RA Peremyslov V.V., Hagiwara Y., Dolja V.V.;  
RT "HSP70 homolog functions in cell-to-cell movement of a plant virus."  
RL Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BV-4;  
RA Peremyslov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;  
RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF190581; AAF14300.1; -;  
DR INTERPRO: IPR000606; -;  
DR INTERPRO: IPR001788; -;  
DR INTERPRO: IPR002588; -;  
DR PFAM: PF00978; RNA\_dep\_RNAPol2; 1.  
DR PFAM: PF01443; Viral\_helicase1; 1.  
DR PFAM: PF01660; Methyltransferase; 1.  
SQ SEQUENCE 3090 AA; 347990 MW; 09C779A57DE874B9 CRC64;

Query Match 6.48; Score 741; DB 12; Length 3090;  
Best local Similarity 22.6%; Pred. No. 6.1e-42;  
Matches 498; Conservative 272; Mismatches 776; Indels 654; Gaps 100;

Qy 408 EYSSV-----VSSITNRLVLRDQALLSHLDTKLCDFMSORDAMIREKPSHRCDVFLKPR 463  
Db 596 EITSLKHKATESYVERVSIHKONLLARSVEKDLIDPKEIKLSKRSRVTVPFYMGEEA 655  
Qy 464 EREKLRELPFELSIOFSDSVSRSHPFANAMRSCFNGIFSRRCGNVCFEDIGGSFTYHKA 523  
Db 656 VOSGLTRAYPOFNLSFTHSVYSDHPAAAGSRLLLENETIASNAKS-SFSDIGCCPLFHKR 714  
Qy 524 GHVCHVNCNPVLDKVKRRNEILFISTAGGDSYVSSDLLTEAASKSVSYCSRESQND 583  
Db 715 GSTDYHVCRTYDKDAQRYSR-ELQARGVLNLSREQULVEAQR-VSVCPHTLGN 771  
Qy 584 SRADAGFMDVYDIPQOVAEAMDKKALVPDIALMPPELVLLYNGEYVLEL--DTLVK 641  
Db 772 VKSDVLIMQVQVYASLNEIASAMVLSKVKVAILMTWTPGELLDEREAFADALGCDVVVD 831  
Qy 642 REGDYLAYNQCCGEMYEHSFSNYSVGF-----FTFSYVRTSSGNVFKLEYEGYRCGVHHL 696  
Db 832 TRDMVOYKFG--SSCYCHKLSLNKIMLTPTFTF-----SGNLFSEVMYENRMGVNY 883  
Qy 697 TMCRAQKSPG-----TEVYRSLVPSPFVSKSLVPIPWAGSSVSFKTIVLDS 743

Db 884 KITRSAYSPEIRGVKTLRYRRACRTEVVQVKL-PRFDKTLTKTFLS-----GYDYIYDA 935  
Qy 744 DFVDRIVSYALNTIGTFENRTFEYAVGAVRQKTHVIGSRVHVKVDISPDMMGLVVA 803  
Db 936 KFSRVDFVYVNSCVSNVSKTFEHWMSYIKSKSRVJISGVHHRDVIDLKHSCFRAV 995  
Qy 804 VMAQAIKDR-----AKSIRSYNFIKASEGLAGVKLF-----FQTVGDCFSNAVSVYAK 853  
Db 996 MLAVGVRSRRTTTEFLAKNLNYTGDAFCFETIRFLFREWSRRAYAEINRSEKLMKSILS 1055  
Qy 854 AMVHDNF-----NVLETLSMPRAPIRKVPVSVVYTICTSGASDRLELRGAFDISKETGR 909  
Db 1056 AGLDYEFLLDLSLQHLLEYSEVEVR-----VSIANGGEVD-----1091  
Qy 910 KLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHC-TRAGLL 968  
Db 1092 --CHNEENRVLTETAE-----AADRK-----SIAAGISGALSSVPTQPRGGLR 1132  
Qy 969 GGSKAT-----VYSSVSKGLVARGAATKAFSGITSEFFSTGSLFYDRGLTDEDERLDALV 1021  
Db 1133 GCSRRSGVSYFLYNLVEEVGNLFESVGDAVFLVKVKFTSDSPIFRVVRMFLD-----LA 1187  
Qy 1022 RTENAINSPVCILETSRVAVS-----KVVAGTKERFWEVSLNDFTTFLVRKNKVLIGI 1073  
Db 1188 EASPFPVSVWSLCAWLREAVSAFSSWADRTVSESVKTFVARTVKRFLNFMSAKTLTKKF 1247  
Qy 1074 FVASLGAAPIA-----WK-YRRGIAANARRYAGSSVETLSLSSSQAAGGLR 1118  
Db 1248 FRFLSASALAKTVVRKAKVILEAYWEVWFESILSDSEY--SAVEFCSSV-----1297  
Qy 1119 GLTSTVSGSLVRRGESSAVTVTR-----ATVAKROVPLALLSFSTSYAISG-----1167  
Db 1298 -ITLLTNSGRLL---PGFSPSAVITEVLLDLATKISIEVLLKQISPVDSSTASSALYRRVL 1353  
Qy 1168 -----CSMLGIW-----AHALPRHLMFF-----FGLGTLGLGAR 1195  
Db 1354 SEILSNFRMTGHEGIFTKVFLLCGLFPLVFRKCVKALCPGDMATYARELEYGVDDDLFLG 1413  
Qy 1196 ASANTWRKFGFSNNMCAPVEVWVRGKSVSLLLPIT-----LCV-SLIIRGLLN-----1243  
Db 1414 RSVNSIK-----NYLCVVAAGLVDSIVDSVVVLKLSGVAKERVGLGFKIILKNFLNPERK 1467  
Qy 1244 -----DTIPQL-----1249  
Db 1468 AKVVTRTSSDLSDEYFSCDESKPQLRGSSRFTLSRLDIFENFLKSKLVNIENACF 1527  
Qy 1250 -AYVPPVEGRNVY-----DETLRYR---DFDY-----DEGAGPSGTQHEAVPGDD 1291  
Db 1528 SAYERIERNMKLYFPPLNSSEEARLIRLCAGDFDYLSDSAFDEDEKLRQAFEQYIYSSDD 1587  
Qy 1292 N-----DGSSTSSVSVVTVNVDYCISTNG-----EVTGEEETHSPRSVQ-YTVVEERVA 1341  
Db 1588 ESVTYDCKPTVLRSYLVNSR-RFLETFCNGPKFPVKVSNYFALYSRLRVLVLPWDRNLS 1646  
Qy 1342 PSAAVAERQGGPSGSGTADAMAFV-----ESYKK-----GYDDVVFH 1377  
Db 1647 DS-----PGLKGNKEKALLAKFLKTCVITACEVSOICCLRLRLCWGTACGLVRLFY 1700  
Qy 1378 QOSGETAREVEVDCKGLLPESVVGEP-----TOEGRADGNTA 1418  
Db 1701 ITYSGTRVLSRVVAVAVCPILLVRNLDGLSDGLTNGMGSVFRRLFVALRRLALSNSA 1760  
Qy 1419 -----QTAV---NEGPREPVQSSSLVSSPOADIPK---VTOSEVHAQ 1454  
Db 1761 LRRKIFEIFGNIIHHPFDVAVIETNEVAPEPL-----SPEVDIDVCCDFGSDSESSESSD 1814  
Qy 1455 EVKOEVLATVSGATPIVDEKPAPSVTTRGVKIIDK--GKA-----VAHVAEKKQ 1502  
Db 1815 EV-ASIPRPG LHGS-----RRSNFLTSLVKVVKLAGIRPLLFLRLRNEVAVYFVER- 1867  
Qy 1503 VOVEQPKORSUTINEGRKAGKOLCMFRFCSCGVQOLDVYNEATIRFSAFTFVNKLKGRS 1562  
Db 1868 -----LASKELKTF-----IGLARLF-----DNFSLTS 1890



```
Db 1248 PRFPLSALAKTVVRKAKVILEAYWEWFESILSDSEY--SAFEFCSSVV----- 1297
Qy 1119 GLTSTVSGSLVRRGFSASVTVTR-----ATVAKQVPLALLSFSTSYAISG----- 1167
Db 1298 -ITLTTNSGRLL-----PGFSPSAVITVLLDLATKISIEVLKQISPVDSTASSALYRVL 1353
Qy 1168 -----CSMLGIW-----AHALPHRLMFF-----FGLGTLICAR 1195
Db 1354 SEILSNRTMGEHIGIFTKVFLLCGLFVFFVKVCVALCPGDMATYARELEYGVDDLEFLG 1413
Qy 1196 ASANTWPKFGFSNMCAVPEVWVRGKSVSSLLLPIT-----LGV-SLIIRGLLN----- 1243
Db 1414 RSVNSIK-----NYLCVAAAGLVDSIVSVWKLKSGVAKERVLGFKSKIIFKNFLNFRK 1467
Qy 1244 -----DTIPOL----- 1249
Db 1468 AKVYRTSSSTDLSEDEYFSCDESKPGILRGSSRFTLSRLDIFNFKLSKSLVIENACF 1527
Qy 1250 -AYVPPVEGRNVY-----DETLRYR---DPDY-----DRGAGPSGTQHEAVPGDD 1291
Db 1528 SAYERIENMKLYFFPLNSSEEAARLIRACAGDFDYLSDSAFDEMLRQAFEQYYSDD 1587
Qy 1292 N-----DGSTSVSSYDVVTVNRVDIGSTNG-----EVTGEEETHSPRSVQ-YTVYVEBEVA 1341
Db 1588 ESVTYDGKPTVLRSLYNVSR-REFLETECNGPKFVKVSNYFKALYSRLRLRVLPMWDRNLS 1646
Qy 1342 PSAAVAERQDPSGSGTADAMAFV-----ESVKK-----GVDDVFF 1377
Db 1647 DS-----PGLKGNERKALLAKIUKTVITACECVSQICCLRLIRLWGTPACGLRFLY 1700
Qy 1378 QQSSGETAREVEVDGKLLPESVVEAP-----TQERGAADGNTA 1418
Db 1701 ITYSGTRVLSRVVAVAVCPLLVRLNELDGLSLTNMGVSVFRLFLVALRALSAYSNSA 1760
Qy 1419 -----OTAV---NEGDRBPVQSSVSSPOADIPK-----VTQSEVHAQK 1454
Db 1761 LRRKIFEFIGNIHHPDVAVETNEVAPEPL-----SPEVDIDVDCDFGSDSESVSSD 1814
Qy 1455 EVKQEVPLATVSGATPIVDEKPAFVTVTRGVKIIDK--GKA-----VAHVAKKKQ 1502
Db 1815 EV-ASIRPRLGHGS-----RRSNFUTSLVKVVKFIAGRIPLRFLRNFVAYVEFERR- 1867
Qy 1503 VQVEQPRQSLITINEGKAGKQLCMFRTCSGCVQDLDVYNEATIAFRFNAETFDVNLKGRS 1562
Db 1868 -----LASRLKTF-----ICLARLF-----DNFSLTS 1890
Qy 1563 AVFF-----SKLG-----EGYTYNGS-----HVSSGWPR-----LEDILTAIKYPSVFD 1603
Db 1891 VVYLLQYDVLNAFIDVELVLNLSGSMVNLPLVS--WVRGSLTKLAEVIVGSGFASFLG 1948
Qy 1604 H--CLVQKYMKGGVPPHADDEECYPSDNPIITLVNLVYKANFSTCKRGGKVMVINVASG 1661
Db 1949 RMCRCVSDW-----CSSLN-----AGCNFMSVPTKKG-FVPPSSSG 1985
Qy 1662 DYFLMPCGFQTHLHSVNSIDEGRISITFRATRRVFGVGRMLQLAGGVSDKEK----- 1713
Db 1986 STAM--YER--LEALESIREHVLSTCR-----VG-----SDEEEERPKVET 2024
Qy 1714 SPGVNPOQSOQATRTITPKSG-----KALSBGSGREVYKGRS 1752
Db 2025 EPGIEHTSEDVPIRSHSQPLSGGECYSSEDRERENRANLLPHYSKIVSERRGLETARN 2084
Qy 1753 TYSIWCEQDYVRKCEWLRADN-----PYMALEPDYTPMTFFVVKGTGTSDEDAVVEYLKY 1805
Db 2085 -----KRTLRGVSEFENALNTSNEQRPRIIV---DHSPS-----RALTVSVREFFYLQE 2131
Qy 1806 LAI-----GIERTYRALIMAR-----NIAVTTAEGVLKVPNVYVESLPGFH- 1846
Db 2132 LALFELSCKLREYYDQLKVFANFNQECCLDKDEDMFVLRAQGVVSGRN---SRPLPKHF 2188
Qy 1847 -----VYKSGTDLIHSTQGLRVLDLPYVLLIAEKGIFTKQKQDVAVVALGDNLFCVDDI 1901
```

```
Db 2189 KDHEFCFRSG-----GL---VPY-----DGTSRVDTI 2212
Qy 1902 LVFHDATNLGALKVARCGMVGSEFKSFEYK-----CYNAPGGGKTTTLVDFEVK-- 1952
Db 2213 --FHTQNFVSA-NALLSGYL--SYRTFTNLNSANVLLYEAPGGGKTTTLIKVFCETF 2267
Qy 1953 SPNSTATITANNGSSEDI-----NMAVKKRDPNLEGLNSATTVNSRVVNFVVRGMYKRLV 2007
Db 2268 SKVNSLILITANKSREELIAKVRNIVLDEGDTPLQTRDRILITIDSYLMN--NRGLTCKVL 2325
Qy 2008 -VDEHMHGQLLQGVFATGASEGLFFGDIQNPFFINREKVFMDCAV-----FVPKK 2060
Db 2326 YLDECFMVHAGAAVACIEFTKCDSAILLFGDSRQIHYIDRNE---LDTAVLSDLNRFV-DD 2381
Qy 2061 ESVVYTSKSYRCPDLVCYLLSMVVRGTEKCYPEKV-----SGKDPVVRSLSKRPDG 2114
Db 2382 ESRVGEVSYRCPMDVCWLSTF-----YPKVTATNLSVAGQS-----SMQVREIE 2428
Qy 2115 TTDDVAEINADVYLCMTQLEKSDM-----KRLGKGGKETPVMTVHEAQKTFSDVVLFR 2169
Db 2429 SVDDVEYSSEFVYLTMLQSEKKDLLKSGFKRSRSEKPT-VLTVHEAQGETYKRVNLVR 2487
Qy 2170 TKKADDSLFTKQPHLVGLSRHTRSLVYAAALSSKLLDDKVGTYISDA 2215
Db 2488 TKFQDDPFRSENHITVALSRHVESLTVSVLSSKRDDAIAQAIVKA 2533
RESULT 6
Q66237 PRELIMINARY; PRT; 3124 AA.
AC Q66237;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE 349-KDA VIRAL POLYPROTEIN.
OS Citrus tristeza virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=12162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T36, QUICK DECLINE;
RX MEDLINE=95266257; PubMed=7747424;
RA Karasev A.V., Boyko V.P., Gowda S., Nikolaeva O.V., Hilf M.E.,
Koonin E.V., Niblett C.L., Cline K., Gumpf D.J., Lee R.F.,
Garnsey S.M., Lewandowski D.J., Dawson W.O.;
RT "Complete sequence of the citrus tristeza virus RNA genome.";
RL Virology 208:511-520(1995).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T36, QUICK DECLINE;
RA Karasev A.V.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U16304; AAC59623.1; -.
DR MEROPS; C42.001; -.
DR MEROPS; C42.002; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Polyprotein.
SQ SEQUENCE 3124 AA; 349350 MW; A02A517D7BDF6C7E CRC64;
```

```
Query Match 6.4%; Score 734; DB 12; Length 3124;
Best Local Similarity 20.5%; Pred. No. 1.9e-41;
Matches 558; Conservative 344; Mismatches 959; Indels 856; Gaps 111;
Qy 47 SDETPRCAAMIIYIGKLTGKVKRTFVPPVK-----GFARQYAVVSGVSALRGD----- 95
Db 594 SNFDELCHAV--LTRMSKGTTLLEILLHSIRERKGLLGAACACFRWSSHFVFKRGNFTARF 651
Qy 96 -----GKKVLMEARTSATSDVSDFDVVFEAVSNALLVVVHVRVVPYAPVKREQ 145
```





```
QY 1208 NMCAVPEVVRGKSVSL--LLPITLGVSLIIR-----GLLN-----D 1244
DB 1196 NNDVLRVIGVFKSLFSLSTLKNLILSPYLKIKWICGKASFTFFVSGLLNVFTQFSD 1255
QY 1245 TIPQALYPPVBERNVYDETLLRYRDFD---YDEGAGSGTQHEAVPGDNDGSGTSSVS- 1300
DB 1256 RINFLVKIALDSEDFQNEVVAAPRNHDLAEFDVTVGVFTKNEFDVVVSSDDFTKKKIISI 1315
QY 1301 -----SYD--VVTVNRDV-----GISTNGEVTGEEETHSPRSVOYTVV 1336
DB 1316 IDELSYERIGIDVQDVQDKKEITELIASSADTPTITKVNHLGHLDDALENDRVRLDILQI 1375
QY 1337 BEEVAPSAVAERQDPSGSGTADAMAFVESV-----KKGVD--DVPH 1377
DB 1376 RKEVL--RWVDDLNEALESGGAGSKLCKDFINYLORNLNLFKHVITSKFGWDFYDILH 1433
QY 1378 QOSSGETAREVEDCGKGLI--PESVVGGAPTQGERGAADGNTAQTAVNEGDEPQVSSILV 1435
DB 1434 -----ESFEVKVRKSAILCAVHDDLSDIYT-----DNLGLVLPKLNKSAREVLYARLI 1481
QY 1436 SSPQADIPKVTQSEVHAQKEVQEVPLATVSGATPIVEKPAPSVT--TRGVKLIIDKGRV 1494
DB 1482 AK-----MKFIYRNEMTTFI-----PSMRFLSGI-----KNKIM 1510
QY 1495 AHV--AEKKOVQVEOPKORSITINEGKAGKQLCMFRTCSGVQOLDVYNEATIAFRSNAF 1552
DB 1511 SHICDAERKQALFKCLDLATLV-----SCATTAOL-----FGGNF 1546
QY 1553 TFDVNLKRSVAVFFSKLGEYTINGSGVSHVSSGWPRALDILTAIKPSVF--DHCLIVOKY 1610
DB 1547 CLIRLI--LSPILSAYLAKSWTYFGEH-----DGIISTIIATSLGNGNYSVNMK 1595
QY 1611 KMGGVPPHADDDEC-----TPSDNPILTVNLVKANFSTKCRK----- 1649
DB 1596 AIVSGSYFHPKPYFCNKAMKNSFRQYVT-DYIARHNEFTTAKTIVGLTSPVGVAVLTV 1654
QY 1650 -----GKVMVIN--VASCD--YFLMPCGFQRTHLHSVNS 1680
DB 1655 GALSINPFAAMFFFAASHVKNFYVDGIVRVSNIAIATGDTLLHLKPVGALKAKINVKR 1714
QY 1681 IDEGRISLTFRATRVFGVGRMLQAGVSEKSPGVPN----- 1719
DB 1715 MKFKNSNITSNT-RVNEVNDGHTAND-TDNKSETSNIAEVEEQSDDDDEIVDIYMK 1772
QY 1720 -----QOPQSQGARTITPKSGKALSEGSGREYKGRSTYSIMCEQDYVRKC 1766
DB 1773 RVKDKMPIKEYIDQSOEPGVKQALT-----NOLLKDSYNDPNASSYSTGA----- 1819
QY 1767 EWL RADNPVMALEPDYTPM-----TFEVVKITGTSADAVVEYKYLAI-GIERTYRALLM 1819
DB 1820 -----VDHLVPDGKYQSVFDIRKNHRGPDISVOLCEELPCEINRNYP----- 1863
QY 1820 ARNIATVTTAEGVLYKPNNOVYESL-----PGPHVYKSGTDLIPH 1857
DB 1864 SKNIVF--QOTVNEVENSIOEVEQVLESLLLNFKKLETTLDKFDGSKMKTROLQTLVND 1921
QY 1858 STQD-----GLVRDLPPVYLIAEKGIPTKGKDVAVALGDNL-FVCDLILVPH 1905
DB 1922 STQFMYVKGWDVWPLSLIHSDPQTV--EYSV-TSSRKLARFSDSGDEVQETTTTEAAGF 1978
QY 1906 DAINLIGALKVARCMGCESEKSEYKCYNAPPGGCKIT-----TLVDEFVKSPTSATIT 1961
DB 1979 SNKKLISMDKFRQSANIIDSRSLSTMKTIVNKPPEGSGKRTQISQMMFDCIIRGKTKCLATS 2038
QY 1962 ANVGSSEDI-----NNAVKKRDPNLEGLNSATTVNSRVNFTVGRMYKRVLVDEVHHMQG 2017
DB 2039 TKVGNELIEKLNLIIVGAKFNKVTQDA---FNMHGKKAIVDELY-----VDECYMSHG 2091
QY 2018 LLQGVFATGASEGLFFGDIQIPFINREKVRMDCAVFPKKESVVYTSKSYRCPDVC 2077
DB 2092 AILMALSNVDFTNCKFFGDENQIPIYSVRLTGTCEALFATTVFKYCHVTHDSISYRCPDVC 2151
```

```
QY 2078 YLLSSMTVRGTEKCYPEKVVSGKDKPVBSLSKRPICGTTDDVAEINADYVYLCMTQLEKSD 2137
DB 2152 YLLSHATDRAGNKMVTNGVYAVGNQN-IRTHILNALTGVANPIREDTVYAAFTQGEKDE 2210
QY 2138 MKRSLKGGKETPVMVTVEHAQGTQFSDVVLFTFKKADDSLFTKQPHILVGLSRRHSLVY 2197
DB 2211 LNRF---AKKNIAKTVNELOGGTGKLNIVFLRLHNNVLYNDTROFITAIRSHRTETFN 2266
QY 2198 AALSCKLDKVGTYTISDASPQSVSDALHTFAPAGC 2233
DB 2267 FCPNSAMNDSISNAINKLN--TIEDNVVAQFGFRQC 2300

RESULT 8
Q08534
ID Q08534 PRELIMINARY; PRT: 3074 AA.
AC Q08534;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.4.8) (CONTAINS: 66 KDA
DE PROTEIN).
OS Sugar beet yellow virus (SBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus
OX NCBI_TaxID=31714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UKRAINIAN (BYV-U);
RX MEDLINE=94082464; PubMed=8259666;
RA Agranovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
RA Lunina N.A., Atabekov J.G.;
RT "Beet yellows closterovirus: complete genome structure and
RT identification of a leader papain-like thiol protease.";
RL Virology 198:311-324(1994).
CC -!- FUNCTION: PROBABLE POLYMERASE.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
CC THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
DR EMBL: X73476; CAA51871.1; -
DR MEROPS: C42.001; -
DR INTERPRO: IPR001788; -
DR INTERPRO: IPR002588; -
DR PFAM: PF00978; RNA_dep_RNAPol2; 1.
DR PFAM: PF01443; Viral_helicase1; 2.
DR PFAM: PF01660; Vmethyltransf; 1.
KW RNA-directed RNA polymerase.
FT CHAIN 1 3074 348 KDA PROTEIN.
FT CHAIN 1 588 66 KDA PROTEIN.
FT MUTAGEN 509 509 C->T: NOT PROCESSED INTO ITS SUBUNITS.
FT MUTAGEN 517 517 C->T: REDUCED PROCESSING INTO SUBUNITS.
FT MUTAGEN 518 518 C->T: REDUCED PROCESSING INTO SUBUNITS.
FT MUTAGEN 556 556 H->E: NO EFFECT.
FT MUTAGEN 569 569 H->E: NOT PROCESSED INTO ITS SUBUNITS.
FT MUTAGEN 588 588 C->D: NOT PROCESSED INTO ITS SUBUNITS.
SQ SEQUENCE 3074 AA; 346028 MW; D75B9564658BE80C CRC64;

Query Match 6.3%; Score 721; DB 12; Length 3074;
Best Local Similarity 21.8%; Pred No. 1.6e-40;
Matches 550; Conservative 318; Mismatches 898; Indels 762; Gaps 117;

QY 137 PYAPVKEQPKPAVKODE-----QKPKQASHWAVKPTA-VGVHVPKPKQKALEPAQSV 190
DB 319 PLRPFREVEIPVVKKHAVPAVVSQVPR-----TRPVATTGAEVYVARNQCSRRPR--- 370
QY 191 PQSLSEKKAALTFGL-----FFSKGGGDESADILRKCKLFN--RALNVPID-----VK 237
DB 371 -NHPILRSASYTFGPKMPLQRFMK---EKREYVVKRSKVSVCSTKSPLEALASILK 425
QY 238 NTFVNA-----KINDEASRRRGYFVVKDRAVKKFPPIVGRATI-----EDFIVNTA- 283
DB 426 NLPOQSYNSERLUKYD-----HFIGDDFEIEVHPLRGKGLUSVLLILPKGEAYCVVTA 478
```







```

Db 2362 -GRCSELDFAVLSDLNRRFV-DDESRVGEVSYRCFWDCAWLSTP-----YFKTVAT 2411
QY 2098 -----SGKDKPVVRSLSKRPGTDDVAEINADVYLCMTQLEKSDM-----KRSUKGKGK 2147
Db 2412 TNLVSAGOS-----SMQVREIESVDDEVSSEFVYLTMLQSEKKDOLLKSFGRKRSRSEK 2466
QY 2148 EPPVMTVHEAQGKTSDDVLFRTKADDSLFTKQPHILVGLSRHTRSLRYAALSCLKDDK 2207
Db 2467 PT-VLTVHEAQGETYRKVNLRKFOEDDPRSENHITVALSRHVESLTYSLSSKRDDA 2525
QY 2208 VGTIISDA 2215
Db 2526 IQAIVKA 2533

RESULT 9
Q9WID7 PRELIMINARY; PRT; 3115 AA.
AC Q9WID7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE P349 PROTEIN.
OS Citrus tristeza virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_Taxid=12162;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99190444; PubMed=10092023;
RA Vives M.C., Rubio L., Lopez C., Navas-Castillo J., Albiach-Marti M.R.,
RT Dawson W.O., Guerri J., Flores R., Moreno P.;
RT "The complete genome sequence of the major component of a mild citrus
RT tristeza virus isolate."
RT J. Gen. Virol. 80:811-816(1999).
RN [2]
RA SEQUENCE FROM N.A.
RA Moreno P.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18420; CAA77161.1;
DR INTERPRO; IPR000515;
DR INTERPRO; IPR000606;
DR INTERPRO; IPR002588;
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Methyltransf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN.1.
SQ SEQUENCE 3115 AA; 346982 MW; 3EEA612F605D5045 CRC64;

Query Match 6.1%; Score 706.5; DB 12; Length 3115;
Best Local Similarity 20.1%; Pred. NO. 1.7e-39;
Matches 482; Conservative 314; Mismatches 826; Indels 777; Gaps 97;

QY 298 MRERAFVCTTKGWCWFFNERLRG-----EYRRRCFSSS-----FSIGFLMHL 340
Db 891 RHIAEVCLSMGKVFRRDVLNAGAPYPYFEVQHRLERYLGRKRAVEYSVRGYFSAPRLYHC 950
QY 341 GFR-SLKVLIRFAGTNILHMPSLNEBERTCGWKGGDYVLNPVKTALVAGDRTLGCILLAS 399
Db 951 CFTDSPRRFRP-----LNKYHKKGGEGALLITDT-----DRLRAGSVYDK 992
QY 400 VANALNQEEVSSVSSITNRLVLRDQSAALLSHLDTKLCDFMSQORDAMIREKPSHRCDFV 459
Db 993 VKKALPD-----SLLVRSVEKDLIEFNQTL-----VDTQ-----RARPGVAVPFO 1032
QY 460 LKPREKRLRELFPPELSQFSDSVSRSSHPFANAMSCENGIFSRRCGNVCFPDIGGSFTY 519
Db 1033 MSENQOVMLTRAYPENINFTSHVSHSDHPVAAGSRMLENHLVRKYAGS-DYSDVGGCPLF 1091
QY 520 HVKACHVNCVNPVLVDKVKRRI-----NEILLFSTAGDSYVSSDLLTEAKSKVS 573
Db 1092 HLRAGTGVHCRPVYDKAHRRVVRHHLQKVAFDHTDGSQV-----GAVNTNT 1143

```

```

QY 574 YCSRSQNCDSRADAGFWVDVYDISPQQVAEAMDKKALVFDIALMFPVELLYNGNEVYL 633
Db 1144 VCGNVLGECFHASEALVMVQVYDPLRELCRAMVNKTSVCYMTMTVPGELLDAESFET 1203
QY 634 BELDLTVKREGDYLAYNYGQC--GEMYEHSFNSVSGFFTFYSVRTS-----SGNVFKLEYE 687
Db 1204 KELLDCSV--ELDVADRVVVCYFNNSAYTHAYSTIR-----EYMRPCVVVDGFLFTIEMV 1256
QY 688 GYRCGYHLLTMCRAOKSPG-TEV-----TYRSLVPSFVGKSLVFIPIVWAGSSVS 735
Db 1257 SLRCSVNYCYITKRCVCPRISEVKRLRYRRCDDLRIKIPRYSSKTRCLP-----G 1309
QY 736 FKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGARSQKTHVITGSRVWISKVDISPD 795
Db 1310 CYLYLDARFVSRYEYVYVNVNVCVNAKTFEWTWNYIKSKSRVVISGKVIHRDVPALIE 1369
QY 796 DMWGLVAVMAQAIKDRAKIRSINYFIKASEGLAGVFKLF-FOYVGDC-----FSNAV 848
Db 1370 YLDGFSAVMLSGAGVGRQNAEAFSRRLAASGD-TSLFELVRFVAVSEKCRDLFVGTQEA 1428
QY 849 SVYKAMVHDNFNVLETLMSPRAFIRKVPGVVVTICTSGASD-RLELRGAFDISKETF 907
Db 1429 ERCVKSRLNSFNMSFVLDSDP-----LLTISEYSELDVPIDLPGGCITEEDE 1477
QY 908 GRKLSRLRVFSRAIVEDSIK---VMKAMKTE-----DGKPLPIT 945
Db 1478 AK-----LIEDGVKSALLRKAVRDEAQLILRGSTSEIVPCGDDDDGLGSDSS 1524
QY 946 EDSVYAFIMGNVSNVHICTRAGLLGSKATVSVSVSKGL-----VAR-----GAA 989
Db 1525 DSSLGS---GSDSSKASSGGLRAGAAPSLSLLKKGCGFCVARRLLDLCLERLVGSL 1581
QY 990 TKAFGS---ITSFFS-----TGSIFYDRLGTEDE--ERLDALVRTENAINSPVILTSRV 1039
Db 1582 SRLKFGTHTLTSTFMSLRKLSNFEDDSIFTAVWDTVMNLIRIVNGVTKTK--LECFRN 1639
QY 1040 A---VSKVVAGTKERFSEVSLNDFTFYLRNKLVIQIFVASLGAAPIAWKYRRIIGIAANR 1096
Db 1640 VLRLIPKLCGCKDF-----AKDFADMC-----KSVLLFST 1670
QY 1097 RVAGSSYETLSSLSQAAG-GLRLGTLSSIVSGSLVVRGFSFSAVTVTRATVAKROVPLA 1155
Db 1671 EFFSKSYTKLCRFSLEIFDVGLEISARN-----GLETCSVSIASAI--DVILT 1718
QY 1156 LLSFSTSYAISGCSMLGIWAHALPRHLMFFGLG--TLIGARASANTWKFSGFSNNWCAV 1213
Db 1719 YPLGSIISFPVVCARVFGALAMEVGSNYCEKTKLGRPEILGEDAFRT----- 1765
QY 1214 PEVWRGKSVSLLLPITLIGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLYRYRDFDY 1273
Db 1766 VSMVLAKAFDKL-----SVDAKGLV-----RCSGIIPGVLRIVFSSLSFS-----DA 1807
QY 1274 DEGAGPSGTQHEAVPGDDNDGSTSSVSSVDVTVNRDVGISTNGEVTGEEETHISPRVQY 1333
Db 1808 DAWGYS--KHE-----VSSLPTI-----TYSKRLSF 1833
QY 1334 TYVEEVAAPSAVAER-----QGDPGSGTADAMAFVESVKKGVDDVDFH 1377
Db 1834 VLEQTKACIAFIVDSSKRLILASLTSEIHESLQNSLTSGAKSLYGRVSIILASNGWYN 1893
QY 1378 QQSSGETAREVEDGKGLLPESVVGGEAPTQERGAADGNTAQTVANNEGPREPVQSSLSVS 1437
Db 1894 K--AGNAVAAVKAAGSG-LSKRVKRIKRPFLHARCVNEQTSSEYYSASDSDSELSDT 1950
QY 1438 POA-----DTPKVTQSEVHAQKEVQVPLATVSGATPIVDEKPAVPTTRGVKIIOK 1490
Db 1951 PGLRNGKCNDLFRVYVSGRVLVYR-----RAVEALK- 1982
QY 1491 GKAVAHVAEKKQOVQEQPKQRSILITINEGKAGKOLCMFRTCSGVOLD--VYNEATATRF 1548
Db 1983 -RLILHVGECLYSRIADSFVRLFVRTNTAVGK--CESLAVDCATKTDLDLVGVYITWRNLH 2039

```

```
Qy 1549 SNAFTFVDNLKGRSAVFSK-LGECYT-----YNGSHVSSGHPRALEDILT 1594
Db 2040 FGAYSTPGLSGGSRVSLRLDGLCTQIONFVTAMLTIFNEFYSLAHRMRTASINV 2099
Qy 1595 AIKYPSPFDHCLVOKYKMGGVPEHADDEECYPSDNPLTILTNLVNKGANFSTCKRK----- 1649
Db 2100 LLEFLKPLPCLVLSIYF-----TECSRIPPL 2127
Qy 1650 -----GKVMYINVASGDYFLMPCGFQ-----RTHLSVNSIDEGRISLTFRAT 1693
Db 2128 KEFLWGFICFVWNIKE-----LSVCYIENLPVCITRLCNRAFRFRSSNEDDERTLQVITEAE 2184
Qy 1694 RRVFCVGHMLQAGVSDSEK-----PCVPNQ--POSQ-----ATRTTPK 1734
Db 2185 NR-----IGFLREVLAEMARYASLEIPVPGSPNPLFPPTYGEIEBELGDGSDSARVIGVD 2241
Qy 1735 SGGKALSE-----GSG----- 1745
Db 2242 NGDSMSTEGEAELSDGSDSDSLPDRDFEQIGLGGSGNVDFRGLRLFRVWRMSNLF 2301
Qy 1746 --REVK-----GRST-----YSI-----W----- 1757
Db 2302 NLRAVKSFGALQOCIFNGRSTGQFISRLFSFGCLPFEPFRHVCVAVRWLSRSLALHVR 2361
Qy 1758 ---CEQDYVRKCEWLADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIIGIERT 1813
Db 2362 NASICDETV---EWLFSFSRFL--YDSRISLVFSEL-----ERVEFYLFSLPKRC 2411
Qy 1814 Y-----RA--LLMARNIAVTAAGVLKVPNOQVYESLPGFHV-YKSGTDLIFHS--- 1858
Db 2412 FTHRSVQVCRAPVRVVRNVCL--VEDGRKSLSEVLNLSLKSVNKKGSAVIESDSS 2469
Qy 1859 -----TODGLRV-----RDLPYVLI 1873
Db 2470 DSESGSGSISFGSINEEVLKTANGSVVSEKSLMKHAAALSGLEPLDGKKDLPIL-- 2527
Qy 1874 AEKGIYTKG----- 1893
Db 2528 --RNFTRGESSGTSVGSLSCEYLTLTNSFGSPDLYPVARDITYKKMTNAMREFYYSQ 2585
Qy 1884 -----DVAVALGD-NLFV----- 1897
Db 2586 KVTLYELHGKLSSYDELKTVGPRKLAKMDLDDVGVDENLKVIVGRHCTRFRFRNTVN 2645
Qy 1898 -----CDDILY-----FHDAINLIGALKVARCGMGE--SFKSFEYK--CYN 1935
Db 2646 SHEPMFCDDGLVLCPPGPKKRDFAISSQTFVAANSFLR-AVDGKDLTFTNSEHSLIYE 2704
Qy 1936 APPGGKTTTLVDFVK--SPNSTATITANVGSSSEDINMAVKKRDPNLEGL-----N 1985
Db 2705 APPGGKTHSLVTVFADYCTKVSLVVTANKNSQAEISQRISREFMDRKLAKHVVKTAG 2764
Qy 1986 SATTVNSRVNFIVRGMYKRVLDVDEHMHIOGLLQLGVEATGASEGLFFGDINOIPFINR 2045
Db 2765 RVFTVDSYLMNLH--RLTDLDFVDECFMVHAGACAVVEFTSCKAVVFGDSROIHIHR 2823
Qy 2046 EKV---FRMDCAVFPVKESVYTSKSYRCPDLVCYLLSSMTVRGTEKCYKPVVSGDK 2102
Db 2824 NDLGVSLLHDIDAFI-EPQHRITYEVSYRCPWDICEWLS-----KEYPRHVATTNG 2874
Qy 2103 PVVR-SLSKRPICGTTDDVAENADVYLCMTQLEKSDMKRSL-----KKGKRETPW-TVHE 2156
Db 2875 SVGKSSVIEVINGCDDVPYDSSAKIIVYTAQEKNEIKLHGLRITVGRTRKAVPVNVTVE 2934
Qy 2157 AOGKTFSDVVLFRTKKADDSLFTKOPHILVGLSRHSTRSLVVAALSSKLDDKVGTVISDA 2215
Db 2935 VQGETYKRVLRVRCYQEDPFCSDNHHVVVALTRHVDLSLTVSVLNSRRYDKTASSIDEA 2993
RESULT 10
Q9IFX0 PRELIMINARY; PRT; 3115 AA.
ID Q9IFX0
AC Q9IFX0;
```

```
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE P349 PROTEIN.
OS Citrus tristeza virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID-12162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T30;
RA Albiach-Marti M.R., Mawassi M., Gowda S., Satynarayana T., Hilf M.E.,
RA Shanker S., Almira E.C., Vives M.C., Lopez C., Guerri J., Flores R.,
RA Moreno P., Garnsey S.M., Dawson W.O.;
RT "Sequences of Citrus tristeza virus separated in time and space are
RT essentially identical.";
RL J. Virol. 74:6856-6865(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T30;
RA Albiach-Marti M.R., Dawson W.O.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF260651; AAF70348.1;
SQ SEQUENCE 3115 AA; 346874 MW; 6B97F494ED25AD63 CRC64;

Query Match 6.0%; Score 689.5; DB 12; Length 3115;
Best Local Similarity 20.8%; Pred. No. 2.7e-38;
Matches 487; Conservative 316; Mismatches 880; Indels 655; Gaps 97;

Qy 298 MRERAFVCTTKGWCWNNERLRG-----EYRRRCFSSS----FSIGFLMHL 340
Db 891 IRHIAEVLCSMGKVFRRDNLGAPPYVFEVQHRLERLYGRAVEYSVRGYSAPRLYHC 950
Qy 341 GFR-SLKVIREAGTINLHMPSLNEERTFGWKGDDVYLPNPVKTALVAGDRTLGLGEILLAS 399
Db 951 CFTDSPRFRF-----LNKYHMGEGEALLITD-----DLRAVGSVYDK 992
Qy 400 VANALNOEEVSSVVSITNRLVLDQDQALLSHLDTKLDMFSDQDAMIREKPSHRCDVF 459
Db 993 VKKALPD---SLVRSVEKDLIEFNQTL---VDTQ-----RARPGVAVPQ 1032
Qy 460 LKPRERKRLRELPELSIQSDSRSRSHPPFANMRSCNGIFSRRCGNVCFFDIGGSTY 519
Db 1033 MSENQOQMLTRAYPEFNINFTSHVSDHPVAAGSRMLNHLVRKYAGS-DYSDVGGCPLF 1091
Qy 520 HVKAGHVNCHVCPVLDVKVKRI-----NEILFLSTAGDSSVSSDLLTEARAKSVS 573
Db 1092 HLRAGHTGIHVCRPVYDVKDAHRVVRHHQLGKVAFDHTDGSQV-----GAVNTNT 1143
Qy 574 YCSRESQCDSDRAGFMVDYVDISPOQVAEAMDKGALVFEDIALMFPELLYNGEVYL 633
Db 1144 VCGNILGCFHASEAIVMVQYVDVPLRELCAWVNKTSVCYMTMVTPELLDARESEFI 1203
Qy 634 EELDTLVKREGDYLAENVGQC--GEMYEHSFNSVSGFTTFYVRTS-----SGNVFKLEYE 687
Db 1204 KELDCSV--ELDVADVADVVCFNNSAYTHAYSTIR-----EYMRTPCVVDGFLTIEMV 1256
Qy 688 GYRCGYHLLTMCRAQSPG-TEV-----TYRSLVPSFVGKSLVFTPVVAGSSVS 735
Db 1257 SLRGSVNYCITKSRVCPRISEVRLRRCDSDLIIRIKIPRYSKTRSCPL-----G 1309
Qy 736 FKTITLSDSDFVDRIYSYALNTIGTFENRTFEYAVGARSQKTHVITGSRVSHKVDISPD 795
Db 1310 CYLYLDAKFFSVRYEYVNNVNVVNAKTFTWNYIKSCSRVVVISCKVLRDVPITALE 1369
Qy 796 DMWGLVAVMAQAATKDRAKSTRSYNFTKASEGSLAGVKLF-FQTVGDC-----FSNAV 848
Db 1370 YLDGFSVAMLSAGYKGRQNAEAFSRRLAASFQD-TSLFELVRFAVSEKRCOLFVGIQEA 1428
Qy 849 SVYAKAMVDHNFVLETLMSPRAFIRKVPGSVVVVTICTSGASDRLELRGAFDLSKETFG 908
Db 1429 ERCVRSYLRNSFNMSFVLDSDP-----LLTI-----SEYSELDVIDL--PGFG 1470
```







```
Db 2247 LDSGDSDDSIPLDRDFEQICGLGCSG-----NVDVFR---LGLKF--VWRMLR 2290
Qy 1557 NLKGRSAVEFS-KLGEVYINGGS-----HVS--GWPRALEDI 1592
Db 2291 NLFNRRAVKSFGALQCIIFNGSGMGFTSRLLSFGNCLPFELFRHVCSVRW--LSKSL 2348
Qy 1593 LTAIKYVSFDHCLVQYKMGCGVFPFHADDECPYSDNPILTVNLVKGANP-----1643
Db 2349 AIRVNRARICDEENVEREFSRILY----DSRYLSD----VFSYLERVDYLSYLSLP 2400
Qy 1644 -----STKCRGGKGVWVTVNASGDYFLMPCQFQRTHLHSVNSIDEGRI-----1686
Db 2401 KRCLTRNRSVRTRAPVRVIR----DMGLVEDG--KKSLETVLNSLDSLKVSDRKGKSAL 2455
Qy 1687 -----SLTERATRVEFGVRMLQ--LAGVSDEKSPGVPNQPOQSGATRT 1730
Db 2456 IESGDSSESCHSGSISFGSVKE-----EVLEAVNGGVSEKAPLMKHADALSR-----2505
Qy 1731 ITPKSGGK-----ALSESGREVKGSTYSIWCEQDYVRKCEWLNRADNPVMALEP 1780
Db 2506 LEPDLSGKKDLPVIRRSFTRGESTGTRPAG-----SICVYLTTLNSSFSGSVP 2552
Qy 1781 DYTPTMTFVVKGTGSDAVVEYLYLAIGIERTYRALLMA-----RNAVTTAEGVLK 1833
Db 2553 DLYPVARDI-----TYKKMTNAMPKPYYSQKVTVLYELHGKL- 2588
Qy 1834 VPNOVYESLP--GFHVYKSGTDLIHFSTQDGLRVDRDLPY-VLIAEKGII--PTKCKDVDAV 1888
Db 2589 --SSYWDLKTGFGNRLKAKMDL-----DDVDVVDVFNKLVIVGRHGTRPPRNTVNSHEF 2641
Qy 1889 VALGDNLFVCD-----DILVPHDAINLIGALKVARGCGWGE--SFKSFEYK--CYNAPPG 1939
Db 2642 MFCDDGLVLCPGPKKRFALISSOTQFVANSFLR-AVDGKDLFTNSEHSLIYEAPG 2700
Qy 1940 GKKTITLIVDEYK--SPNSTATITANVGSSEDIINAVKKRDPNLEGL-----NSATT 1989
Db 2701 GKCKTSLVTVFARYGCTKVSLCVVTVANKNCAEISORISRELMDRKMLAKHVVKTAGRVFT 2760
Qy 1990 VNSRVAFIVRGMYKRVLDVDEHMHGGLQLQGVFATGASGLFPGDINOIPIFNREKV- 2048
Db 2761 VDSLMMHI--RLKTDLLFVDVCFWVHAGAIGAVVEFTRKAVVFGDSRQIHYIHRNDFG 2819
Qy 2049 --FRMCAVFPKKSVMYTSKSRCLPDLVCYLLSSMTVRGTERCYPEKVVSGKDKPVVR 2106
Db 2820 VSLLDHIDAFI-EPQHRIGYGVSYRCPWDICWLS-----KFYPRHVATANTGSIGR 2870
Qy 2107 -SLSKRPIGTTDDVAEINADVYLCMTOLEKSDMKRSLGK-----GKETPW-TVHEAQG 2159
Db 2871 SSVSIEILINGDDVPYDSSAKYIVYTOAEKNELOKHL--GRITVAGTRAVPVTNVHEVQG 2929
Qy 2160 KTFSDVVLFRTKKADDSILFTKQPHILVGLSHRHSLSVYALSSALSDDKVGTIISDASPOS 2219
Db 2930 EYTKKRLVIRCKYQEDTPSSCDNHHVVVAL/HRVDSLTYSVLNSRRYDKTASSIDEA--RE 2987
Qy 2220 VSDALLHTEAPAGC 2233
Db 2988 IFDKFRSTNHSHGC 3001

RESULT 13
Q83044 PRELIMINARY; PRT; 1873 AA.
AC Q83044:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE METHYLTRANSFERASE.
OS Lettuce infectious yellows virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus
OX NCBI_TaxID-31713;
RN [J]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-92;
RA Klaassen V.A., Boeshore M., Koonin E.V., Tongvan T., Falk B.W.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U13440; AAA61797.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf_1.
DR Transferase; Methyltransferase.
SQ SEQUENCE 1873 AA; 217264 MW; 55A77D8357761F0 CRC64;

Query Match 5.6%; Score 643; DB 12; Length 1873;
Best Local Similarity 18.4%; Pred. No. 1.8e-35;
Matches 406; Conservative 323; Mismatches 733; Indels 748; Gaps 83;

Qy 166 VKPRAVGVHVPPLPKQEALEPAQSPVQOSLEEKAAALTFGLFFSKGGGSDAVILRKGL 225
Db 259 VKPRVSVN-----KAREVSSSVKQLQYREGEV-----FVYSKACKE 298
Qy 226 FNRALNPIDVKNTFVWAKIWDASRRRGYFYVKDRAVKFPPIVGRATIEDFIVNTAPG 285
Db 299 DTRVLNDDNAIRNLF-----NATLNGGKY-----KLHP-----326
Qy 286 CDVALPRIELMSMRERAFVCTTKGCWCFN-----NERLURGETYRRRCFSSFSIGFLMHL 340
Db 327 -----EASATGKRFYKDFGFCWLDVFADANRRI--PEWVKPHCLLTG---SVLMSC 374
Qy 341 G---FRSLKVIIRFAGTNILHMPSLNEERTFGWKGDVYLPVNPKTAIVAGDTRILGGEIL 397
Db 375 GLWDFAKKMYVS--VSHGLLH-----YDRKLESSARAGVRDFVG---412
Qy 398 ASVANALNOEVEYSV---VSSITNRLV----LRDOSALLSHLDTKLDCMFESQORDAMTRE 450
Db 413 --ASNEAVQREDFDLDLRVEEFABRVLETANLRSDNRLDINILTRASDYINKKSKESE 470
Qy 451 KPSHRCDVFLKPREKRELPELSIQFSDSVSSSHHPFANAMKSCFNGIFSRRCGNVCF 510
Db 471 LDINVC---LSMDEKKMITNLFDPDIOMSFNOKSYNHCYFNAMRACENFYFSRKFKNDSY 527
Qy 511 FDIGGSFTYHVKAGHVNCHVCPVLDVKDKRRRINEILFLSTAGGDSYVSSDLLEAAASK 570
Db 528 IDAGGVSTVTRKSNHNHVCSPRLDLKDAARHIQIATVIDLKG-----YGE 575
Qy 571 SVSYCSRESQNCDSRADAGFMVDYDTSPOQVAEMDKKALVFDIALMFPVELLYNGGE 630
Db 576 TISCTNKTEDCAVNRDIIIAVEYDMLTMDMAKAMLSHGSRKFECFIIPEIFTKECN 635
Qy 631 VYLEELDTLVKRGEDLAYNVGQCGEMYEHFSFNVSGFTEFYVRTSSGNVFKLEYEGR 690
Db 636 VELYEGRLKVTIGDGVVEYYGSGNETFSHSCQTLKDILSVQ--VFQGGRVFKKLTLEHSR 694
Qy 691 CGYHHLTMCRAQK--SPQTEVTYRSLVPSFVGKSLVFIIPVAVGSSVSFKTIYL--DSDFVDR 748
Db 695 GOLHFFSTICEKIEPGSVKLTYYQSRSELDKVTLRIPVKDSFGVWVTHYIILKEDREFVSS 754
Qy 749 IYSVALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVHVSKVDISPDMDWGLVAVMAQA 808
Db 755 MIEYVANTGIKIDDKMVEWTYSQYRAKKTVTIKSGKVTKQETRIKRELIPGFIATIMSEG 814
Qy 809 IKDRAKSIRSYNFIKASEGS-----LAGVFKLFFQTVGDCFSNAVSVYAKA 854
Db 815 IRAREKTHYLAKMLYTSHYKYSIYNIIFRLIMHFLGGTKRFIYESLVDCLEFLT-----868
Qy 855 MVHDNFNVLETMSMPRAFIRKVPQSVVVVITCTSGASDRDLRGAFDISKETFGKRLKNS 914
Db 869 ----NSDYIDITVN-----TESRIE-----DLDK-----888
Qy 915 RLRFVSRVAVEDSTSVMKAMKMTEDGKPLP-ITEDSVYAFIMGNVSNVHCTRAGLLGSKA 973
Db 889 -----WPFVQPNVTI-----TTDAEDQPSILEQSVKFTL-----917
```









Db 173 RIKRQADEIKIYENLKNKRAQPEDLDGKWYCGNKFEECVFRAERAYAMAIHSIYDIDL 232

Qy 601 QVAEAMDKGALVFDIALMFPVELLYNGEYVLEELDLTKREGDYLAY-----NVGOC 654

Db 233 DLANALEEKRIKIMSGTFLFSVDLLGKSGSLPTMDFFFEYEGDYVKYGFKNNDTPG-- 290

Qy 655 GEMYHSPNSVSGFTTFSSVRTSSGNVFKLEYEGVRCGYHLLTMCRAQKSPGFEVYRSL 714

Db 291 ---YKHKLSQLMKYLTKFVAKGGTIYYLELTOEGDVMFTMTDA-----TEAMNGV 342

Qy 715 V--PSFV-----GKSLVPIPVV-----AGSSVFKTIVLDSFVDRIYSYA----- 753

Db 343 VADESFKCIPIDNKSEVFPFIEVDKQDLSLIFSEVLLSKDPVQRAIETGRKLPNQLTP 402

Qy 754 --LNTIGFENRTFAYGAVRSQTHV-----ITSRVHSHKVDISPDDMMGL----- 800

Db 403 DNVNTYLTSTNNIIIG--GSSKNNTKVDATLIQIOTTLLVYMT-----ELMNLQRK 454

Qy 801 VVAVMAQAIDKRA--KSIRSYNFIKASEGSLAGVFKLFOTVGDGCFNSAVSVYAKAMVHD 858

Db 455 VLEKRLQMKDDVDFKSLAHTAFLKFV-GKVS-----YTORALRCFANWLS-----YVH- 502

Qy 859 NFNVLETLMSMPRAFIRKVPVSVVVTICTSGASDRLELRGAFDISKETFGRLKNSRLRV 918

Db 503 -----CTDALQPHD-----V 512

Qy 919 FSRRAIVEDSIKVMKAMKEDGKPLPITEDSVYAFIMGNVSN-----VHCTRAGLLGSGK 972

Db 513 PLYAEVTDVRLKWKTHAPNOGVFLDM-EDLDVKIKMHVESREKRDVSRICVSGKLGEIQ 571

Qy 973 ATPVSSVSKGLVARGAATKAFSGITSTFSGSLFYDRGLTDERLDALVRTENAINSPVG 1032

Db 572 IHSESS-----NNEEQINGDYRDSRRRTTFED 598

Qy 1033 ILETSRVAVSKYVAGTKFWSVSLNDETFVLRNKVLIGIFVASLGAAPIAWKYRRGTA 1092

Db 599 LLE-----GEVATNFDLDWCE--KNDHFN----- 621

Qy 1093 ANARYAGSSYETLSLSQAAGLUGLTSSTVSGSLVVRGRFSVAVTVTRATVAKROV 1152

Db 622 --SRDDAOSKY-----AWGLKLLK----- 638

Qy 1153 PLALLSFSTSYAISGCSMLGIWAHALPHLMFFFGGLTLLGARASANTWKFGFSNNWCA 1212

Db 639 -----GIWEFLPP--MDF----- 650

Qy 1213 VPEVVMRGKSVSLLPLITGLSVLIIRGLNDTIPQLAYVPPVEGRNVYDETLRYRDFD 1272

Db 651 -----APV-----YVDDEQARLRNRVNERNTSD----- 675

Qy 1273 YDEGAGPSGTQHEAVPGDDNDGSTSVSS-----YDVVTNV--RDVGISTNGEVTG 1321

Db 676 -----TGADGDVAACEVAADFESKAMDTLVDVYVKMDQKRLGL----- 712

Qy 1322 BEETHSPRSVOYTYVEEVEAPSAVAERQDPSGCTADAMAFVESVKGVDDVDFHQOSS 1381

Db 713 -----VEKVKASAVAVELSKTKPVGTNSCAVDLWADFENKLD-----DDE 754

Qy 1382 GETAREVEVDGKGLLPESVGEAPTQERGAADGNTAGTAVNEGDPREPVOSSLVSPQAD 1441

Db 755 PCCGISVYTKKEICQDNV---GPV-----LLCGSSSCSSVSEVEKE-----TD 795

Qy 1442 IPKVTOSEVHAQKEVQEVPLATVSGATPIVDEKPAVSTTRGVKIIDKRAVAHAEKK 1501

Db 796 VVSVTD-----SGNSP-----ATCCGMVLDSVSI-----GN 824

Qy 1502 QVOVEOPKORS--LTINSGKACKQLCMFTCSGVQLDLYNEATATRFSNAFTVDNLK 1559

Db 825 DVSVELPVTATVATVETGTA-----DGVIASTDESAFSDVND-- 862

Qy 1560 GRSVAFSKLGEYTYNGSHVSSGWRPALEDILTAKYPSVDFHC--LVQYKMGGVGP 1617

Db 863 -----KSWGVAEB-----ESDFTFYSCGLISDKVKKSNLP 894

Qy 1618 FHADDEECPSPNDPILTVNLVGNKANFSTKCRKCGKVMVINVASGDYFLMPCGQRTHLHS 1677

Db 895 KRPDF-----SKFPTVQOKAKHEAMWYLOCK-----IVS-----DRTTLRS 930

Qy 1678 VNSIDGRISLTFRATRRRVFGVGRMLQLAGGVSDKSPGVNPOQPOSGQATRTITPKSGG 1737

Db 931 I--IDDLHLCMFHNGTCEL-----PKDSAFLDYTDNCGTWMYKGPSPRFG 973

Qy 1738 KALSEGSGREVKGRSTYSIWCEQDYVRKCEWLNRADPNVMALEPDYTPMTFEVYKVTGTSED 1797

Db 974 HSVGVGFSLDTQR-----VSKC-----ELVKLAWNHD 1001

Qy 1798 AVVEYLKYLAIGIERTYRALLMARNIAVTAEGVLKVPNQVYESLPGFHVYKSGTDLIFH 1857

Db 1002 S-----RGQV----- 1006

Qy 1858 STQDGLRVRLDPLVYLAEKGIETKGDVDVAVVALGDNLFVCDLILVPHDAINLIGALKVA 1917

Db 1007 -NOKPVNTRAFQYLLUSELS-----FMNEMIITYR--NIQQVMKRK 1044

Qy 1918 RCGMVGESEKSPKCYNAPPGGKTTTLVDEFEVKSPNSTATITANVG--SSEDIINMAVK 1975

Db 1045 -----ERSKOARITLRDGVPCGCKSTWILN---NANPTKOMILSVGKEATEDLKEKPM 1094

Qy 1976 KRDPNLEG--LNSATTVNSRVNPIVRGMYKRVLVDEVHMHQGLLQLGVPATGASGLFF 2034

Db 1095 KKHRCVESDLKRITRVDSFLMHDYDKYRAATVHFDEALMAHAGMVFCAIDILGAKKVICQ 1154

Qy 2035 GDINOIPIFNR-EKVFPMDCAVFPKESVVTYSKSYRCPLDVCYLLSSMTVVRGTEKCP 2093

Db 1155 GDSQQIPFINRVESITLQYSKLVIDETEHVRLT---YRSPVDVAHYLTK-----KSWYS 1205

Qy 2094 EKVWSGDKPVVRSLSK-----RPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGGK 2147

Db 1206 GGRVITKN--PVMRSKMTVGPVROVKPMTSVHCVPYFKDAQYLTFTQSEKTDLYKALRNKGP 1264

Qy 2148 ETPVMTVHEAOGKTFSDVVLFRTKKADDSLF-----TKQPHILVGLSRHRSRLVYALSSK 2203

Db 1265 VT-VNTVHETOGKTDFDVIIVRLETTENEIYPGGRKGQPYEIVATTTHRRSLVY---YTA 1320

Qy 2204 LDDKVCTYISDASPOSVSDA--LLHTFAP 2230

Db 1321 IEDRLFEDISDM--QDVMESKLMKSLCP 1346

Search completed: January 31, 2001, 15:35:03  
Job Time: 548 sec

**This Page Blank (uspto)**